

256700-1-F

Final Report

WAVELET TRANSFORM BASED TARGET DETECTION

N. S. Subotic, L. Collins, M. Reiley, B. Thelen, J. Gorman

MAY 1995

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

Prepared for:
Army Missile Command
Redstone Arsenal, AL

19960201 025

Contract No. DAAH01-93-C-R346

DTIC QUALITY INSPECTED 5



ERIM

P.O. Box 134001
Ann Arbor, MI 48113-4001

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1284, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave Blank)		2. REPORT DATE May 1995		3. REPORT TYPE AND DATES COVERED Final Report
4. TITLE AND SUBTITLE Wavelet Transform Based Target Detection			5. FUNDING NUMBERS	
6. AUTHOR(S) N. Subotic, L. Collins, M. Reiley, B. Thelen, J. Gorman				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) ERIM P.O. Box 134001, Ann Arbor, MI 48113-4001			8. PERFORMING ORGANIZATION REPORT NUMBER 256700-1-F	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Army Missile Command Redstone Arsenal, AL			10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION/AVAILABILITY STATEMENT To protect information not owned by the U.S. Government and protected by a contractor's "limited rights" statement, or received with the understanding that it not be routinely transmitted outside the U.S. Government.			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) A novel method of target detection based on a multiple resolution representation of SAR images is described. The signature variation due to interference between scatterers is exploited as resolution is varied. Prominent target scatterers interfere in a characteristic manner as resolution is changed; background noise (clutter) will remain random. Statistical models characterizing the behavior of the multiresolution signature processes for both clutter and cultural objects are derived. A number of detection strategies that exploit these processes are formulated. The detection strategies were applied to simulated SAR scenes containing extended, complex targets in a homogeneous clutter background. Significant detection improvement was demonstrated over pixel-by-pixel CFAR methods using only the fine resolution signature. These detection strategies were also applied to collected ERIM DCS and Lincoln Laboratory ADTS data. We found that using multiresolution based detection strategies provide a significant increase in detectability of targets.				
14. SUBJECT TERMS Multiresolution Processing, Synthetic Aperture Radar, Target Detection			15. NUMBER OF PAGES 114	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UNLIMITED	

Abstract

A novel method of target detection based on a multiple resolution representation of SAR images is described. The signature variation due to interference between scatterers is exploited as resolution is varied. Prominent target scatterers interfere in a characteristic manner as resolution is changed; background noise (clutter) will remain random. Statistical models characterizing the behavior of the multiresolution signature processes for both clutter and cultural objects are derived. A number of detection strategies that exploit these processes are formulated. The detection strategies were applied to simulated SAR scenes containing extended, complex targets in a homogeneous clutter background. Significant detection improvement was demonstrated over pixel-by-pixel CFAR methods using only the fine resolution signature. These detection strategies were also applied to collected ERIM DCS and Lincoln Laboratory ADTS data. We found that using multiresolution based detection strategies provide a significant increase in detectability of targets.

Contents

1	Introduction	1
2	Multiresolution Process Statistics	7
2.1	Case 1: Natural Clutter	7
3	Detection Strategies	14
3.1	GLRT for Multiresolution AR Processes	14
3.1.1	Special Cases	17
3.2	Composite Tests	20
3.3	Generalized Matched Filter	21
4	STAR Algorithm Description	25
4.1	Introduction	25
4.2	Algorithm Description	27
4.2.1	Contrast Enhancement	27
4.2.2	CFAR Screening	30
4.2.3	Aggregation	31
5	Evaluation of Multiresolution CFAR Performance	33
5.1	Synthetic Scene Generation	33
5.2	Empirical Results – Simulated Data	36
5.3	Empirical Results – STAR Data	54
5.4	Results - Discrimination	56
5.4.1	Feature Extraction	56
5.4.2	Discrimination Algorithm/Results	62
6	Conclusions	66
	Appendix A Resolution Sampling Strategy	68
A.1	First Sampling Strategy	73
A.2	General Sampling Strategy	77
A.3	Solution to Resolution Sampling Strategy	80
A.4	Summary	93
	Appendix B Properties of Noisy Wavelet Transform	95
	Appendix C Reproducing Kernel Hilbert Spaces Representation	104
C.1	Karhunen-Loeve Expansion.	105
C.2	Reproducing Kernel Hilbert Spaces	106

List of Figures

1	Top: Simple three-scatterer target example. Middle: Magnitude signature of simple three-scatterer target as a function of resolution. Bottom: Phase signature of simple three-scatterer target as a function of resolution.	5
2	Magnitude signature of random clutter as a function of resolution. . .	6
3	Pictorial representation of a rect aperture weighting showing that the scaling law in correlation is satisfied.	10
4	Sampling strategy for the choice of resolutions to whiten the clutter increments process in resolution.	12
5	Target and clutter regions which are tested in the GLRT for similarity.	15
6	Receiver operating characteristics for the multiresolution detectors using the signal based on three point scatterer target.	23
7	Original STAR multi-channel contrast-based detection algorithm flow.	26
8	Algorithm flow of multiresolution/STAR detection algorithm.	28
9	Portion of the data set used in the detector performance studies showing embedded targets with a 0 dB target-to-clutter ratio.	35
10	Detection statistics for various single pixel detection strategies.	38
11	ROC results for various single pixel multiresolution detection schemes.	39
12	Detection statistics for the various multiresolution/spatial strategies.	41
13	ROC performance of the various spatial multiresolution test strategies.	42
14	Histograms of the decision statistics for the various strategies.	43
15	ROC performance of the multiresolution F test as the number of resolutions are increased. Starting resolution is 1ft.	44
16	ROC performance of the multiresolution F test as the number of resolutions are increased. Starting resolution is 1.5ft.	45
17	ROC performance of the multiresolution P test as the number of resolutions are increased. Starting resolution is 1ft.	46
18	ROC performance of the multiresolution P test as the number of resolutions are increased. Starting resolution is 1.5 ft.	47
19	ROC performance of the multiresolution F test as the starting resolution is coarsened.	48
20	ROC performance of the multiresolution P test as the starting resolution is coarsened.	49
21	ROC performance of the multiresolution tests for a starting resolution of 1.5 ft.	50
22	ROC performance of the multiresolution tests for a starting resolution of 2 ft.	51

23	ROC performance of the multiresolution tests for a starting resolution of 3 ft.	52
24	ROC performance of the multiresolution tests for a starting resolution of 5 ft.	53
25	Target to clutter ratio gain afforded by multiresolution processing. . .	57
26	ROC performance of the multiresolution tests applied to the STAR data set.	58
27	ROC performance of the multiresolution tests applied to the DCS data set.	59
28	ROC performance of the single resolution and multiresolution AR tests when the starting resolution is coarsened.	60
29	Detection/discrimination algorithm flow of multiresolution/STAR algorithm.	61
30	Results of the discrimination algorithm on both the single resolution and multiresolution data.	65

List of Tables

1	List of the various multiresolution detection schemes used in the performance studies.	36
2	List of features extracted for each aggregate blob.	61

1 Introduction

In this report we describe a target detection concept which uses the change of the complex-valued synthetic aperture radar (SAR) signature in a pixel when resolution is varied. The conjecture is that there is a specific signature change between groups of cultural scatterers which is different from the signature change in random clutter when resolution is varied. This change in signature is due to the interference of the prominent scatterers as they contribute to the signature in the resolution cell. As resolution changes, different target prominent scatterers enter/exit a resolution cell. Clutter, on the other hand, will have large numbers of equivalued scatterers causing it to remain random with respect to resolution.

Many synthetic aperture radar (SAR) target detection algorithms work on a single resolution image at the finest available resolution [20], [21]. However, there is compelling evidence to suggest that significant performance gains can be achieved by casting the detection problem in a multiresolution setting. We argue, using simple physical principles, that this performance gain is a direct result of the coherent interference effects that occur in typical radar target signatures as the imaging resolution is varied. We then apply a multiresolution sampling strategy for choosing optimal resolutions for coherent target detection.

SAR images of man-made objects typically consist of spatial patterns of bright points and lines resulting from radar backscatter from discrete physical features such as corners, edges, flat plates and other primitive geometric shapes. The coherent radar return from each of these discrete features, or *prominent scatterers*, is a complex phasor with amplitude equal to the local radar cross-section of the target feature. At fine enough resolutions, these prominent scatterers are isolated in individual resolution cells and they dominate the target signature. As the resolution changes from fine to coarse, adjacent scatterers become lumped together into a single resolution cell and coherently interfere with each other, leading to characteristic changes in amplitude and phase as a function of resolution. It is this relationship between

phase and amplitude as a function of resolution that we exploit through the use of a multiresolution-based detector.

There are two potential benefits which can be gained from exploiting this discriminant. The first benefit is that more reliable target detection (higher probability of detection for a given false alarm rate) can be obtained. The second is that the required aperture for equivalent target detection performance to a non-multiresolution case may be reduced and therefore, the system search rate is increased. These benefits were empirically verified. An increase in detection probability of approximately a factor of 2 was achieved for a given false alarm rate using multiresolution signatures. In addition, the use of multiple resolutions attained an equivalent performance to a pixel-by-pixel fine resolution approach without the need for the fine resolution data.

In our development of SAR imaging, we will assume linear FM transmission signals and the usual far-field assumptions [2]. Neglecting polar formatting concerns (data is assumed interpolated or backprojection processing is used) and higher order phase effects (range walk and variable range rate), the complex image can be expressed in terms of the scene, $s(\underline{x})$, $\underline{x} \in \mathbb{R}^2$, as

$$T(\underline{x}; \rho) = \frac{1}{\rho} \int \int s(\underline{y}) e^{i2\pi \left[\frac{|\underline{y}|^2}{\lambda r_a} \right]} h\left(\frac{\underline{x} - \underline{y}}{\rho}\right) d\underline{y} \quad (1)$$

$$= \frac{1}{\rho} \int \int c(\underline{y}) h\left(\frac{\underline{x} - \underline{y}}{\rho}\right) d\underline{y} \quad (2)$$

The quadratic phase term $e^{i2\pi \left[\frac{|\underline{y}|^2}{\lambda r_a} \right]}$ arises as a result of retaining the second order term in the binomial series expansion of the distance from the SAR antenna to a specific point in the scene. This quadratic term incorporates the interference (phase mixing) of the various scatterers within a resolution cell. This effect is usually accounted for by defining a complex scene $c(\underline{x})$. As resolution is varied the scatterer interference will take on a unique characteristic determined by the strength and relative location of the scatterers within the resolution cell.

To obtain an intuitive understanding of the role that interference of prominent

scatterers plays a simple example is now presented. The example is depicted at the top of Figure 1. A set of 3 scatterers are equidistantly placed in azimuth. The scene can then be written as

$$s(\underline{x}) = \sum_{i=1}^3 \delta(\underline{x} - \underline{x}_i) \quad (3)$$

For an unweighted phase history, the complex image of this scene is

$$T(\underline{x}; \rho) = \sum_{i=1}^3 e^{\frac{i2\pi|\underline{x}_i|^2}{\lambda r_a}} \text{sinc}\left(\frac{\underline{x} - \underline{x}_i}{\rho}\right). \quad (4)$$

The magnitude and phase of this target signature for 1, 2, and 3 scatterers are shown in the middle and bottom of Figure 1, respectively. These figures show the effect of interference between the scatterers when they enter a resolution cell. In the case of a single scatterer, no interference is present and no change in signature is detected as a function of resolution. For 2 and 3 scatterers the signature changes significantly. When each individual scatterer can be resolved, there is no interference between scatterers (neglecting effects from higher order sidelobes). As the resolution degrades, the first order sidelobe has an effect on the signature of the scatterer. When the resolution becomes large enough to contain all of the scatterers, the change in signature is most pronounced.

Natural terrain typically consists of a large collection of small amplitude scatterers that are randomly distributed within each resolution cell. These random paths are shown in Figure 2. Thus SAR imagery of terrain, i.e., *clutter*, is frequently modeled as a Gaussian random field by appealing to the law of large numbers [23] and references therein. The result is that the amplitude and phase of a clutter pixel vary randomly as a function of resolution. In particular, for Gaussian clutter the amplitude has a Rayleigh distribution with parameter proportional to the resolution and clutter reflectivity, while the phase is uniformly distributed over $[0, 2\pi)$.

The resolutions where the multiresolution discriminant is applicable must be coarse enough to allow multiple prominent scatterers in a resolution cell but not so coarse that the law of large numbers can be applied for target signatures. As

shown in [1], approximately 8 scatterers of equal value in a resolution cell is the limit where the signature can be said to be nonrandom. The choice of resolutions in which the multiresolution feature vector is constructed must be chosen appropriately.

A natural concern that arises when images of various resolutions are to be jointly processed/analyzed is the correspondence of a pixel from image to image. This problem can be circumvented by choosing the maximum aperture extent (finest resolution) a priori, sampling at a rate defined by the Nyquist rate at that resolution and retaining that sampling rate as the resolution changes. The images at coarser resolutions are effectively oversampled. Each pixel in each of the multiresolution images will now have a direct correspondence with each other.

In section 2 we will develop the statistical multiresolution process models. This section will also discuss a simple resolution sampling strategy which whitens the multiresolution clutter process. Section 3 will discuss the detection strategies that we will employ in our studies. The detection strategies will be based on a composite hypothesis test which determines the presence of a mean value in the process. Special cases encompassing a Generalized Likelihood Ratio (GLRT) approach will also be discussed. A second resolution sampling strategy based on a generalized matched filter is also developed and discussed in the appendices. Section 4 discusses the overall detection algorithm developed in the Strategic Target Algorithm Research (STAR) [22] program where our detection strategies were inserted. Section 5 discusses the performance results of our detection strategies when applied to simulated data. Section 5 also discusses the detection results when the multiresolution based algorithm was applied to the Lincoln Laboratories STAR data set and to ERIM DCS data. A set of appendices are also given which review a multiresolution sampling strategy when a generalized matched filter strategy is used and some properties of noisy wavelet transforms.

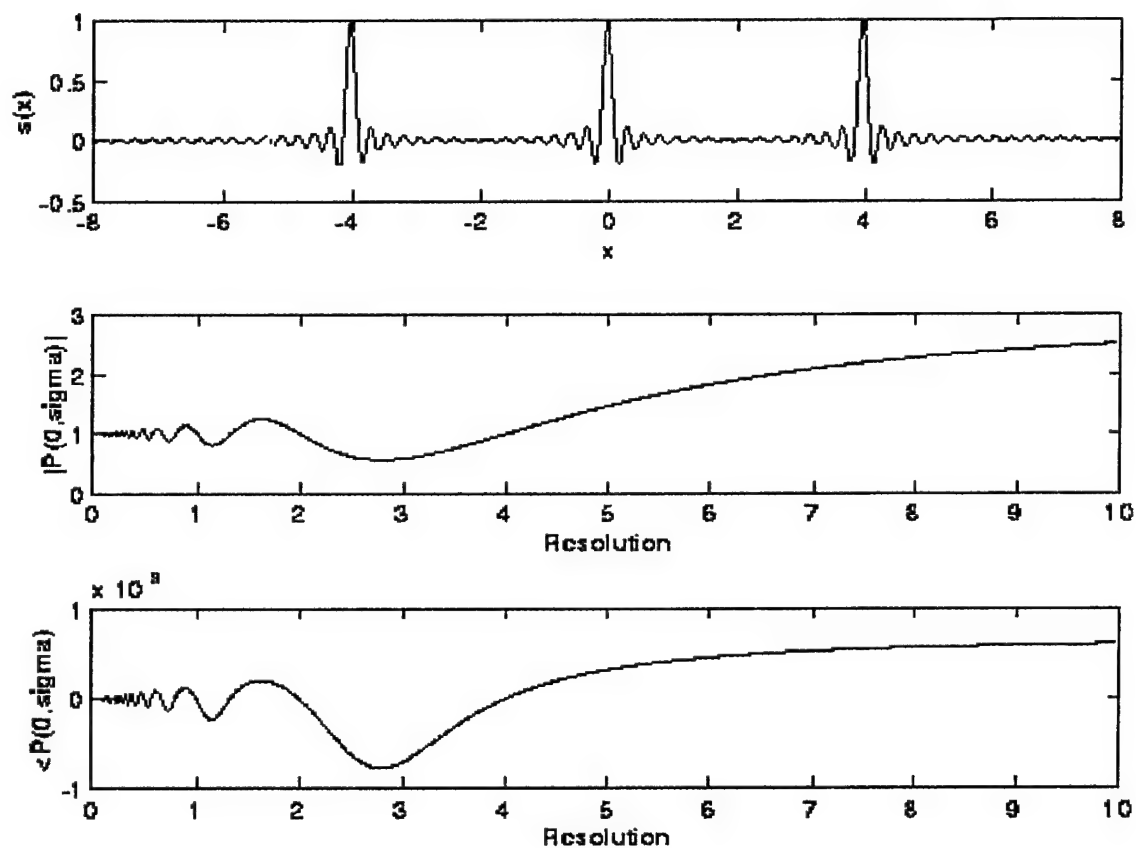


Figure 1: Top: Simple three-scatterer target example. Middle: Magnitude signature of simple three-scatterer target as a function of resolution. Bottom: Phase signature of simple three-scatterer target as a function of resolution.

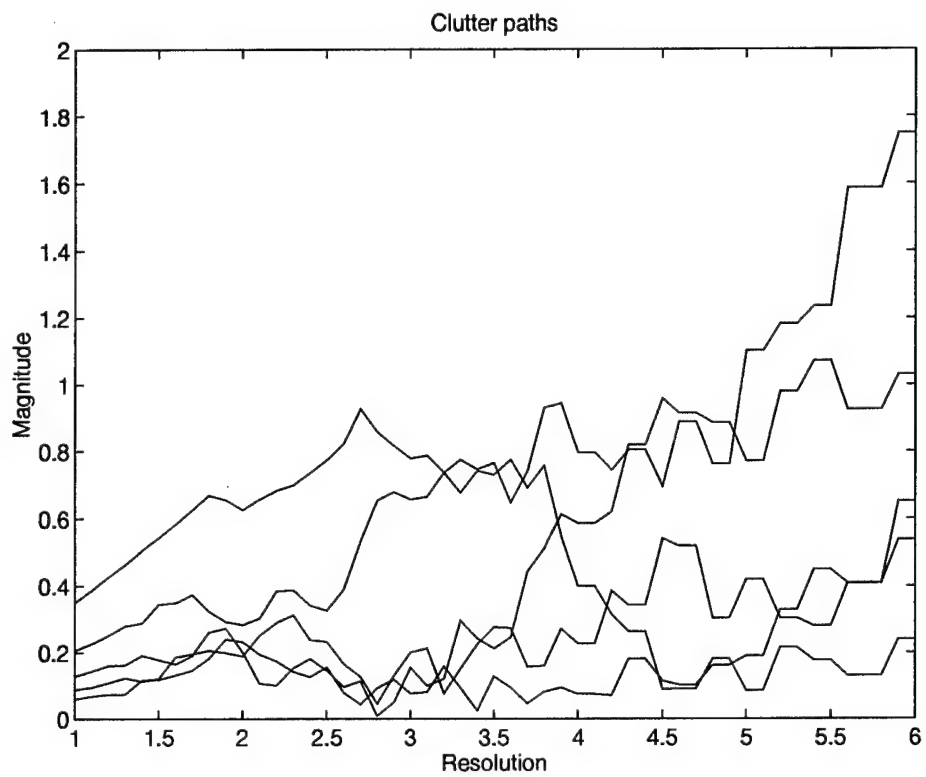


Figure 2: Magnitude signature of random clutter as a function of resolution.

2 Multiresolution Process Statistics

In our construction, we model the scene $c(\underline{x})$, $\underline{x} \in \mathbb{R}^2$ as a collection of point scatterers, in which each point scatterer is specified by its location $\underline{x}_k \in \mathbb{R}^2$ and complex reflectivity $u_k \in \mathbb{C}$:

$$c(\underline{x}) = \sum_{k=1}^K u_k \delta(\underline{x} - \underline{x}_k). \quad (5)$$

This approach has been used to model both clutter [23, and references therein] and objects that consist of collections of point reflectors [19, for instance] (i.e., trihedrals or corner reflectors). We refer to c as the complex reflectivity function.

The complex-valued SAR image, $T(\underline{x}; \rho)$, taking into account resolution, can be written as a convolution between the complex reflectivity function, $c(\underline{x})$, and the system impulse response, $h(\underline{x})$, [2]:

$$\begin{aligned} T(\underline{x}; \rho) &= \frac{1}{\rho} \int c(\underline{y}) h\left(\frac{\underline{x} - \underline{y}}{\rho}\right) d\underline{y} \\ &= \frac{1}{\rho} \sum_{k=1}^K u_k h\left(\frac{\underline{x} - \underline{x}_k}{\rho}\right). \end{aligned} \quad (6)$$

If we model the locations $\{\underline{x}_k\}$ as a Poisson point process with intensity $\lambda(\underline{x})$, then $c(\underline{x})$ is a compound point process or marked Poisson process with marks $\{u_k\}$ [24, Chapter 3]. $T(\underline{x}; \rho)$ forms a filtered Poisson process at each resolution ρ [24, Chapter 4].

In the next two subsections, we consider models for the statistics of the observed radar image, T , under simple assumptions on the statistics of the scatterer locations $\{\underline{x}_k\}$ and their complex reflectivities $\{u_k\}$.

2.1 Case 1: Natural Clutter

For natural terrain, i.e., *clutter*, a typical assumption is that each resolution cell in the SAR image contains a large number of small amplitude scatterers. In this case, we will assume that the complex reflectivities $\{u_k\}$ are independent, identically distributed

(*iid*) random variables with mean zero and covariance $\sigma_c^2 I$ where I is the identity matrix.

Using a slight generalization of a result in [23], we can invoke the Generalized Multivariate Central Limit Theorem [16] to show that the joint density of $T(\underline{x}; \rho)$ converges to multivariate Gaussian in both space and resolution as the number of scatterers K tends to infinity. For a set of *iid* Circular Complex r.v. X_1, \dots, X_N , a set of arbitrary weights a_{j1}, \dots, a_{jN} $j = 1, \dots, J$ where

$$\left(\frac{1}{\sqrt{\sum_{k=1}^N a_{1k}^2}} \sum_{k=1}^N a_{1k} X_k, \dots, \frac{1}{\sqrt{\sum_{k=1}^N a_{Jk}^2}} \sum_{k=1}^N a_{Jk} X_k \right) \sim \mathcal{N}_c(0, \Sigma) \quad (7)$$

By assuming that the clutter scene $c(\underline{x})$, $\underline{x} \in \mathbb{R}^2$ consists of a set of random point scatterers ($c(\underline{x}) = \sum_k u_k \delta(\underline{x} - \underline{x}_k)$) with u_k *iid*, zero mean and with covariance $\sigma_c^2 I$, the summation of the scatterers within a resolution cell produces to complex gaussian random variable. The weights described in equation 7 correspond to the impulse response of the system.

The mean and covariance of the process are

$$E\{T(\underline{x}; \rho)\} = 0, \quad (8)$$

and

$$\begin{aligned} \Sigma(\underline{x}, \underline{x}'; \rho, \rho') &= R(\underline{x}, \underline{x}'; \rho, \rho') \\ &= E\{T(\underline{x}; \rho) T^*(\underline{x}'; \rho')\} \\ &= \frac{\sigma_c^2}{\rho \rho'} \int h\left(\frac{\underline{x} - \underline{y}'}{\rho}\right) h^*\left(\frac{\underline{x}' - \underline{y}'}{\rho'}\right) dy', \end{aligned} \quad (9)$$

where $R(\underline{x}, \underline{x}'; \rho, \rho')$ is the correlation function of $T(\underline{x}; \rho)$, and σ_c is the variance of the *iid* complex reflectivities $\{u_k\}$. Note that the covariance function Σ is completely specified by the variance σ_c and the impulse response $h(\underline{x})$.

The choice of the impulse response h influences the statistics of the SAR image $T(\underline{x}; \rho)$. This is especially true when examining $T(\underline{x}; \rho)$ as a function of resolution. For a specific spatial location $\underline{x} = \underline{x}' = \underline{x}_j = \underline{0}$, the covariance in resolution becomes

$$\begin{aligned}\Sigma_{\underline{x}_j}(\rho, \rho') &= \Sigma(\underline{0}, \underline{0}; \rho, \rho') \\ &= \frac{\sigma_c^2}{\rho\rho'} \int h\left(\frac{-\underline{y}}{\rho}\right) h^*\left(\frac{-\underline{y}}{\rho'}\right) d\underline{y}.\end{aligned}\quad (10)$$

If the impulse response is chosen so that the correlation function R satisfies a scaling law condition [5] with respect to resolution, then the process $T(\underline{x}; \rho)$ is Gauss-Markov in resolution. For $\rho_l \leq \rho \leq \rho_u$ the scaling law is

$$R_{\underline{x}_j}(\rho_l, \rho_u) = \frac{R_{\underline{x}_j}(\rho_l, \rho) R_{\underline{x}_j}(\rho, \rho_u)}{R_{\underline{x}_j}(\rho, \rho)}.\quad (11)$$

where $R_{\underline{x}_j}(\rho_l, \rho_u) = R(\underline{x}_o, \underline{x}_o; \rho, \rho')$. The scaling law as it pertains to the SAR impulse response becomes

$$\int h\left(\frac{-\underline{y}}{\rho_l}\right) h^*\left(\frac{-\underline{y}}{\rho_u}\right) d\underline{y} = \frac{1}{\|h\|^2} \int h\left(\frac{-\underline{y}}{\rho_l}\right) h^*\left(\frac{-\underline{y}}{\rho}\right) d\underline{y} \int h\left(\frac{-\underline{y}'}{\rho}\right) h^*\left(\frac{-\underline{y}'}{\rho_u}\right) d\underline{y}'.\quad (12)$$

where $\|h\|^2 = \int |h\left(\frac{-\underline{y}}{\rho}\right)|^2 d\underline{y}$.

The simplest means of analyzing equation 11 is through the Fourier transform of the system impulse response

$$h(\underline{x}) = \int H(\underline{f}) e^{i2\pi(\underline{f} \cdot \underline{x})} d\underline{f}.\quad (13)$$

The scaling law can now be written in terms of the aperture weighting as

$$\int H(\rho_l \underline{f}) H^*(\rho_u \underline{f}) d\underline{f} = \frac{\int H(\rho_l \underline{f}) H^*(\rho \underline{f}) d\underline{f} \int H(\rho \underline{f}') H^*(\rho_u \underline{f}') d\underline{f}'}{\|H\|^2}\quad (14)$$

For $h(\underline{x}) = \text{rect}(\underline{x})$ or $H(\underline{f}) = \text{rect}(\underline{f})$ and $\rho \leq \rho'$, the correlation of the process in resolution becomes

$$\begin{aligned}R_{\underline{x}_o}(\rho, \rho') &= \rho' \int H\left(\frac{\rho' \underline{f}}{\rho}\right) H^*(\underline{f}) d\underline{f} \\ &= 4\rho'.\end{aligned}\quad (15)$$

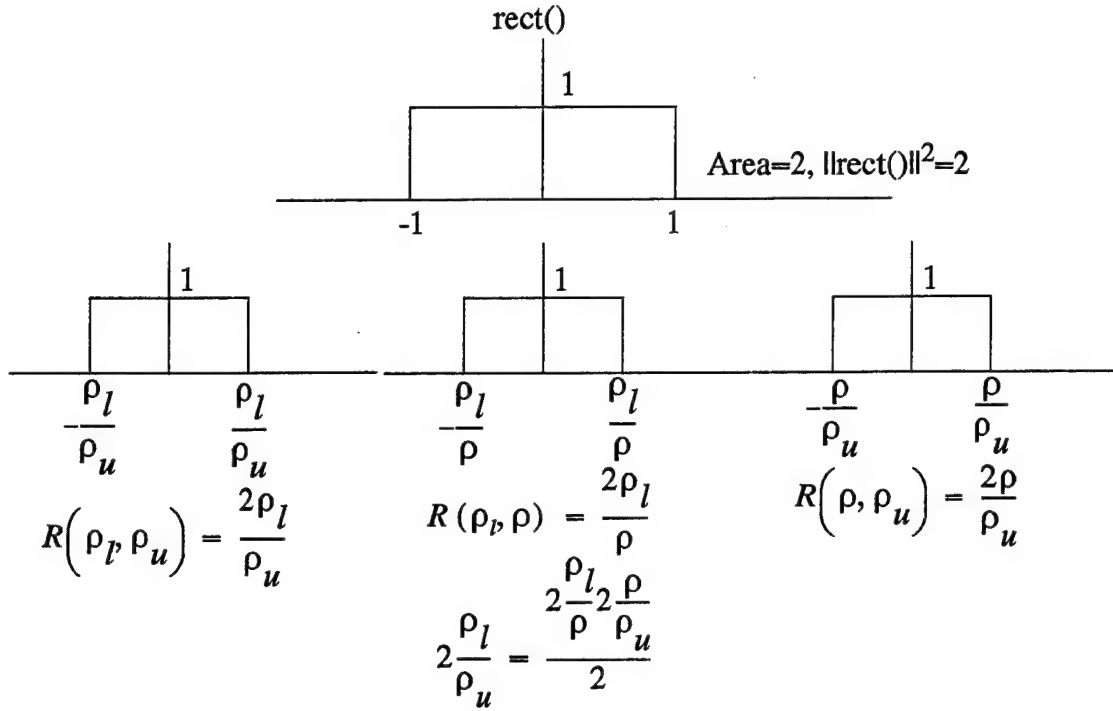


Figure 3: Pictorial representation of a rect aperture weighting showing that the scaling law in correlation is satisfied.

A simple substitution of equation 15 into equation 11 shows that the scaling law is satisfied when the system impulse response is either $\text{sinc}(\underline{x})$ or $\text{rect}(\underline{x})$ and the process is Gauss-Markov. This is simply shown pictorially in Figure 3. The correlation of $T(\underline{x}_j; \rho)$ becomes $R_{\underline{x}_j}(\rho, \rho') = \sigma_c^2 / \max\{\rho, \rho'\}$ with variance $\sigma^2 = \sigma_c^2 / \rho$.

$T(\underline{x}; \rho)$ can also be shown to have independent increments in resolution - i.e.

$$E\{[T(\underline{x}_j; \rho_1) - T(\underline{x}_j; \rho_2)][T(\underline{x}_j; \rho_2) - T(\underline{x}_j; \rho_3)]^*\} = 0 \quad (16)$$

for $\rho_1 > \rho_2 > \rho_3$. In this case, the clutter process $T(\underline{x}_j; \rho)$ is a Brownian motion process when viewed as a function of $\frac{1}{\rho}$ [7].

The Brownian motion nature of T in resolution can be exploited to provide a simple linear transformation of the process which whitens the process in resolution. This transformation is based on the independent increments of the resolution process and the scaling of the variance in resolution. Choose a set of resolutions $\rho_1 < \dots <$

$\rho_i < \rho_{i+1} < \dots < \rho_N$ where $\rho_{i+1} = \rho_i + \delta\rho_i$. An increments process in resolution is formed by

$$T'(\underline{x}_j; \rho_i) = T(\underline{x}_j; \rho_i + \delta\rho_i) - T(\underline{x}_j; \rho_i) \quad (17)$$

with $\delta\rho_i > 0$. This process has zero mean and is independent from resolution to resolution.

Furthermore, by judicious choice of resolutions $\{\rho_i\}$, the variance of the difference process T' can be made constant from resolution to resolution. We choose adjacent resolutions to satisfy $\frac{1}{\rho_i} - \frac{1}{\rho_{i+1}} = \gamma$ for every i . Under this latter condition, the resolution step size $\delta\rho_i$ can be easily shown to be

$$\delta\rho_i = \frac{\gamma\rho_i^2}{1 - \gamma\rho_i}. \quad (18)$$

This resolution sampling strategy is shown in Figure 4 for two choices of γ . The resolution sampling is dense at fine resolutions and becomes sparse at coarse resolutions.

For a fixed point \underline{x}_j , we define the vector of resolution increments

$$\underline{T}'(\underline{x}_j) = \{T'(\underline{x}_j; \rho_1), \dots, T'(\underline{x}_j; \rho_{N-1})\}^t \quad (19)$$

where t denotes vector transpose. Then $\underline{T}'(\underline{x}_j)$ has distribution $\underline{T}'(\underline{x}_j) \sim \mathcal{N}_c(\underline{0}, \gamma\sigma_c^2 I)$ where $\mathcal{N}_c(\underline{\mu}, \underline{\Sigma})$ denotes a circular complex Gaussian density with mean $\underline{\mu}$ and complex covariance $\underline{\Sigma}$. Here the symbol \sim is shorthand notation for the phrase “has probability distribution.” We also define the vector

$$\underline{T}(\underline{x}_j) = \{T(\underline{x}_j; \rho_1), \dots, T(\underline{x}_j; \rho_N)\}^t \quad (20)$$

which is a vector of samples from the original multiresolution process at the same set of resolutions as the increments process.

Case 2: Statistics of Cultural Objects

Many man-made or cultural objects typically consist of a small number of large amplitude point scatterers. In the case of our multiresolution analysis, the physical

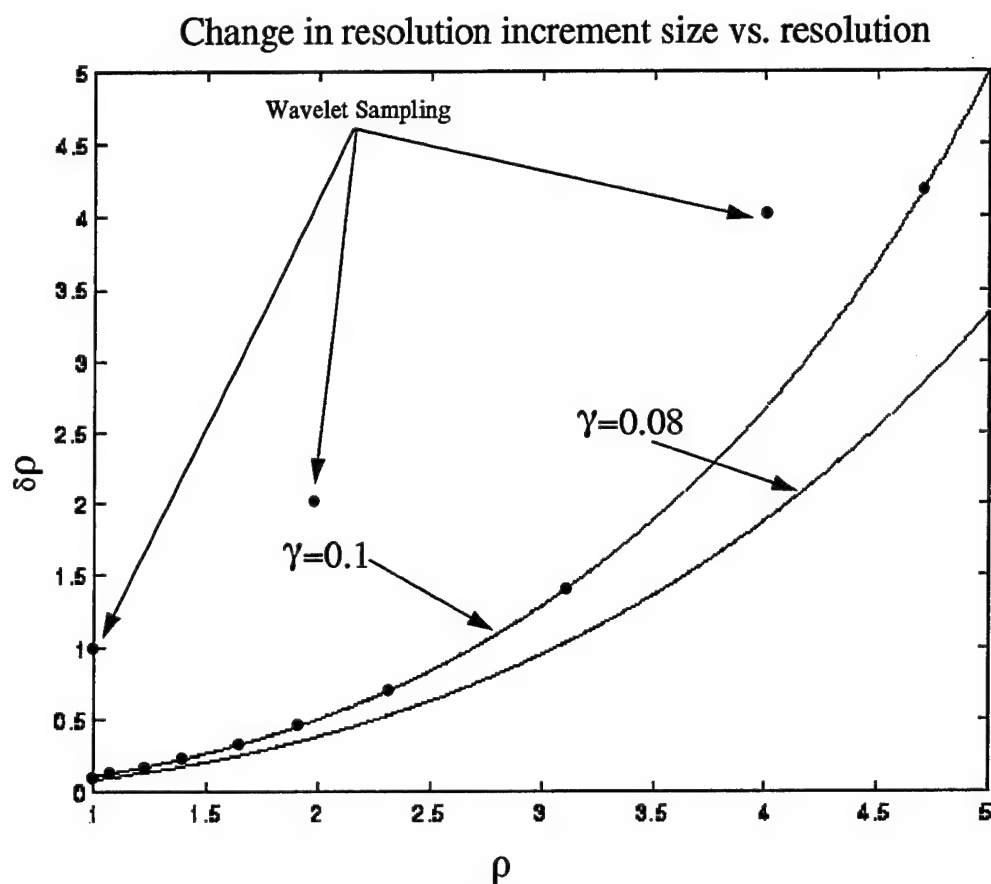


Figure 4: Sampling strategy for the choice of resolutions to whiten the clutter increments process in resolution.

phenomena we wish to explore is the interaction of a small number of local prominent scatterers and the interference patterns that result. The number of local prominent scatterers we wish to examine are typically less than eight. Larger numbers tend to make the SAR signature exhibit zero mean complex Gaussian statistics [1].

We will reformulate the process to have a random and nonrandom component. The point scatterer model will become

$$c(\underline{x}) = \sum_{k=1}^K u_k \delta(\underline{x} - \underline{x}_k) + \sum_{k=1}^{K'} a_k \delta(\underline{x} - \underline{x}_k). \quad (21)$$

The first term in Equation 21 will correspond to the clutter model as outlined in Case 1. The second term will correspond to an unknown set of prominent scatterers. The complex reflectivities $\{a_k\}$ are deterministic but unknown. The spatial distribution of the prominent scatterers is also assumed to be deterministic but unknown. $T(\underline{x}; \rho)$ will be multivariate Gaussian with covariance as in Case 1 (Eq. 10). The mean of the process is

$$E\{T(\underline{x}; \rho)\} = \frac{1}{\rho} \sum_{k=1}^{K'} a_k h\left(\frac{\underline{x} - \underline{x}_k}{\rho}\right). \quad (22)$$

In the same spirit as Case 1, we form a vector of increments in resolution, $\underline{T}'(\underline{x}_j)$. This process is complex Gaussian with distribution $\underline{T}'(\underline{x}_j) \sim \mathcal{N}_c(\underline{\mu}, \gamma \sigma_c^2 I)$ where $\underline{\mu} = \{\mu_1, \dots, \mu_i, \dots, \mu_{N-1}\}^t$ and

$$\mu_i = E\{T'(\underline{x}_j; \rho_i)\} = \sum_{k=1}^{K'} a_k \left[\frac{1}{\rho_i + \delta \rho_i} \left(\frac{\underline{x}_j - \underline{x}_k}{\rho_i + \delta \rho_i} \right) - \frac{1}{\rho_i} h\left(\frac{\underline{x}_j - \underline{x}_k}{\rho_i}\right) \right]. \quad (23)$$

3 Detection Strategies

The goal of our detection scheme is to construct simple pixel screening algorithms. Our detection strategies will exploit information provided by resolution. This information is, in fact, produced by the local spatial signature as it is incorporated into the resolution process. We derive three classes of multiresolution tests. The first class is the Generalized Likelihood ratio test (GLRT). This test exploits the oscillatory behavior of the multiresolution signature due to interference between scatterers. We use an autoregressive (AR) signal model to characterize the oscillatory behavior. We derive the general multiresolution GLRT using these signal models and a set of simplified tests based on special cases of the models. The second class of tests is composite hypothesis tests whose purpose is to detect differences in the mean value of the multiresolution process. As shown in Section 2, when faced with a deterministic but unknown set of scatterers (cultural object), there is a mean value on the multiresolution process whereas clutter will be a zero mean process. Lastly, we will briefly explore the construction of a generalized matched filter detector. This detector formulation assumes that the multiresolution signature is known *a priori*. We use the generalized matched filter as a means of deriving a resolution sampling strategy.

3.1 GLRT for Multiresolution AR Processes

With the GLRT testing strategy, we are interested in testing for large local changes in the statistical behavior of the data. Specifically, our comparison is between the statistics of a local area \mathcal{X}_t presumed to contain a target and another area \mathcal{X}_c presumed to be clutter. The configuration of the two areas under test is shown in Figure 5. Due to the oscillatory nature of the multiresolution signatures as shown in Figure 1 we will use an autoregressive (AR) model for the multiresolution signatures. Each area will have its own first order AR process model in resolution described as

$$T(\underline{x}_j; \rho_i) = a_c T(\underline{x}_j; \rho_{i-1}) + c_i \quad \underline{x}_j \in \mathcal{X}_c \quad (24)$$

$$T(\underline{x}_j; \rho_i) = a_t T(\underline{x}_j; \rho_{i-1}) + t_i \quad \underline{x}_j \in \mathcal{X}_t \quad (25)$$

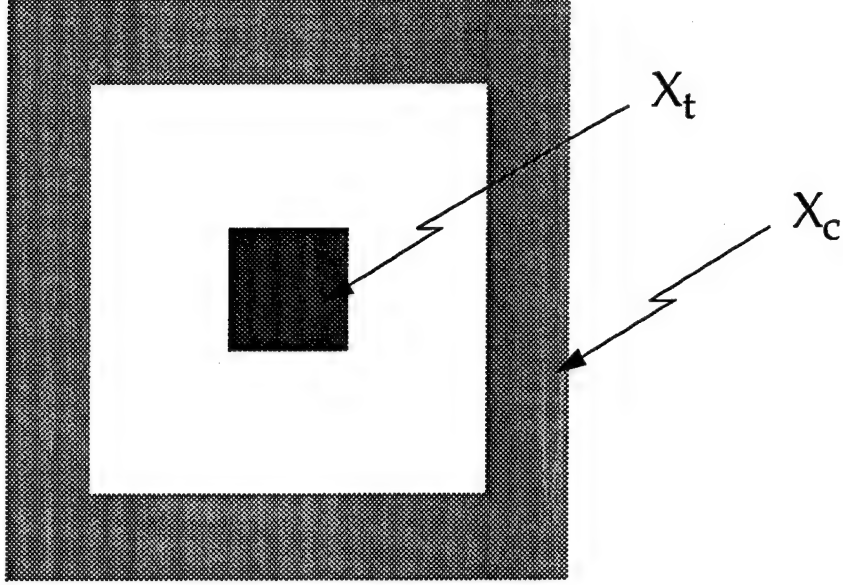


Figure 5: Target and clutter regions which are tested in the GLRT for similarity.

where c_i and t_i are zero mean Gaussian random variables with variances σ_c^2 , and σ_t^2 respectively. We are assuming that the AR processes are spatially independent and identically distributed over each local area.

We wish to test whether the inner area conforms to the same AR process as the outer clutter area. The hypotheses under test are

$$\begin{aligned} H_0 : & \quad a_c = a_t, \quad \sigma_c^2 = \sigma_t^2 \\ H_1 : & \quad a_c \neq a_t, \text{ or } \sigma_c^2 \neq \sigma_t^2 \end{aligned}$$

Our approach is to construct a GLRT to test for the difference between the two areas. Let us define a data vector \underline{T}_l , $l = t, c$ which incorporates the data in each local area as

$$\underline{T}_l = \{T(\underline{x}_1; \rho_1), \dots, T(\underline{x}_1; \rho_N), T(\underline{x}_2; \rho_1), \dots, T(\underline{x}_2; \rho_N),$$

$$\dots, T(\underline{x}_l; \rho_1), \dots, T(\underline{x}_l; \rho_N)\}^t \quad \underline{x}_j \in \mathcal{X}_l, \quad l = t, c. \quad (26)$$

Our test will use N resolutions, J_c pixels in area \mathcal{X}_c and J_t pixels in area \mathcal{X}_t . The joint distribution for the spatial/multiresolution process in \mathcal{X}_l is

$$p_l(\underline{T}_l) = \left(\frac{1}{\sqrt{2\pi}\sigma_l} \right)^{NJ_l} \exp \left[-\frac{\sum_{\underline{x}_j \in \mathcal{X}_l} \sum_{i=1}^N |T(\underline{x}_j; \rho_i) - a_l T(\underline{x}_j; \rho_{i-1})|^2}{2\sigma_l^2} \right] \\ \cdot \exp \left[-\frac{\sum_{\underline{x}_j \in \mathcal{X}_l} |T(\underline{x}_j; \rho_0)|^2}{2\sigma_l^2} \right] \quad l = t, c. \quad (27)$$

The GLRT is defined as [4]

$$\Lambda(\underline{T}) = \frac{\max_{\theta_1} p_1(\underline{T}|\theta_1)}{\max_{\theta_0} p_0(\underline{T}|\theta_0)} \stackrel{H_1}{>} \beta \quad (28)$$

where \underline{T} is the entire data set encompassing both regions. $\theta_k = [a_k, \sigma_k^2]^t$, $k = 0, 1$ are the unknown parameters in our test for each hypothesis. The extrema of this test are satisfied using the maximum likelihood estimates of the parameters conditioned on the hypotheses [4]. We will denote the parameter vector corresponding to these estimates as $\hat{\theta}_k = [\hat{a}_k, \hat{\sigma}_k^2]^t$, $k = 0, 1$.

To evaluate the GLRT, we will assume that the data in the two regions are independent. Recall that under hypothesis H_1 the parameters of the AR processes are assumed to be different. Using the assumption that the statistics of \mathcal{X}_c and \mathcal{X}_t are independent, the extrema of the joint conditional distribution $p_1(\underline{T}|\theta_1)$ can be decomposed as $\max_{\theta_1} p_1(\underline{T}|\theta_1) = p(\underline{T}_c|\hat{\theta}_c)p(\underline{T}_t|\hat{\theta}_t)$. The parameters for each area are estimated separately. Let $\hat{\theta}_l = [\hat{a}_l, \hat{\sigma}_l^2]^t$, $l = t, c$ be the parameter vectors corresponding to the maximum likelihood estimates based solely on the areas \mathcal{X}_l . These parameter estimates are given as

$$\hat{a}_l = \frac{\sum_{\underline{x}_j \in \mathcal{X}_l} \sum_{i=1}^N T(\underline{x}_j; \rho_i) T^*(\underline{x}_j; \rho_{i-1})}{\sum_{\underline{x}_j \in \mathcal{X}_l} \sum_{i=1}^N |T(\underline{x}_j; \rho_{i-1})|^2} \quad (29)$$

$$\hat{\sigma}_l^2 = \frac{1}{NJ_l} \sum_{\underline{x}_j \in \mathcal{X}_l} \sum_{i=1}^N |T(\underline{x}_j; \rho_i) - \hat{a}_l T(\underline{x}_j; \rho_{i-1})|^2, \quad l = t, c \quad (30)$$

Under hypothesis H_0 , the AR models are assumed to be identical. Consequently, the unknown coefficients are estimated using both regions \mathcal{X}_c and \mathcal{X}_t . Again, using the assumption that the statistics of \mathcal{X}_c and \mathcal{X}_t are independent, the extrema of the joint conditional distribution $p_0(\underline{T}|\theta_0)$ can be decomposed as $\max_{\theta_0} p_0(\underline{T}|\theta_0) = p(\underline{T}_c|\hat{\theta}_0)p(\underline{T}_t|\hat{\theta}_0)$. Let $\hat{\theta}_0 = [\hat{a}_0, \hat{\sigma}_0^2]^t$, and Let $\mathcal{X} = \mathcal{X}_c \cup \mathcal{X}_t$. These estimates are

$$\hat{a}_0 = \frac{\sum_{\underline{x}_j \in \mathcal{X}} \sum_{i=1}^N T(\underline{x}_j; \rho_i) T^*(\underline{x}_j; \rho_{i-1})}{\sum_{\underline{x}_j \in \mathcal{X}} \sum_{i=1}^N |T(\underline{x}_j; \rho_{i-1})|^2} \quad (31)$$

$$\hat{\sigma}_0^2 = \frac{1}{NJ_c J_t} \sum_{\underline{x}_j \in \mathcal{X}} \sum_{i=1}^N |T(\underline{x}_j; \rho_i) - \hat{a}_l T(\underline{x}_j; \rho_{i-1})|^2 \quad (32)$$

For the case where many more clutter pixels are used than target $J_c \gg J_t$, $\hat{a}_0 \simeq \hat{a}_c$, $\hat{\sigma}_0^2 \simeq \hat{\sigma}_c^2$, and $\hat{\theta}_0 \simeq \hat{\theta}_c$. The GLRT can then be approximated as

$$\Lambda(\underline{T}) = \frac{\max_{\theta_1} p_1(\underline{T}|\theta_1)}{\max_{\theta_0} p_0(\underline{T}|\theta_0)} = \frac{p(\underline{T}_c|\hat{\theta}_c)p(\underline{T}_t|\hat{\theta}_t)}{p(\underline{T}_c|\hat{\theta}_0)p(\underline{T}_t|\hat{\theta}_0)} \simeq \frac{p(\underline{T}_t|\hat{\theta}_t)}{p(\underline{T}_t|\hat{\theta}_c)} \stackrel{H_1}{>} \beta. \quad (33)$$

The log GLRT is then given as

$$\begin{aligned} \psi(\underline{T}) = \log \Lambda(\underline{T}) &= NJ_t (\log(\sqrt{2\pi}\hat{\sigma}_c) - \log(\sqrt{2\pi}\hat{\sigma}_t)) \\ &+ \frac{1}{2\hat{\sigma}_c^2} \sum_{\underline{x}_j \in \mathcal{X}_t} \sum_{i=1}^N |T(\underline{x}_j; \rho_i) - \hat{a}_c T(\underline{x}_j; \rho_{i-1})|^2 \stackrel{H_1}{>} \beta \end{aligned} \quad (34)$$

To help mitigate the assumption of statistical homogeneity over the target area we will only label the center pixel of the area \mathcal{X}_t with the outcome of the test and not the entire target area which would be standard. This creates a pixel-by-pixel test where the local target area is used to help provide averaging in the estimation of the AR coefficient and variance.

3.1.1 Special Cases

The special cases that we examine are motivated by models for the process statistics of the clutter. For natural terrain, a typical assumption is that each resolution cell

in the SAR image contains a large number of small amplitude scatterers. We will assume that the complex reflectivities of the scatterers are independent, identically distributed (*iid*) random variables with mean zero and covariance $\sigma_c^2 I$ where I is the identity matrix.

Special Case 1: AR vs. Brownian motion

This clutter model provides a motivation to test between an AR process indicative of a target vs. Brownian motion clutter. Brownian motion is simply an AR process where $a_c = 1$ [25]. The GLRT then simplifies to the test

$$\begin{aligned} \psi_5(\underline{T}) = & NJ_t(\log(\sqrt{2\pi}\hat{\sigma}_c) - \log(\sqrt{2\pi}\hat{\sigma}_t)) \\ & + \frac{1}{2\hat{\sigma}_c^2} \sum_{\underline{x}_j \in \mathcal{X}_t} \sum_{i=1}^N |T(\underline{x}_j; \rho_i) - T(\underline{x}_j; \rho_{i-1})|^2 \stackrel{H_1}{>} \beta \end{aligned} \quad (35)$$

We can create a simpler test by assuming $\sigma_c^2 = \sigma_t^2$. If we also examine the case where $\mathcal{X}_t = \underline{x}_j$ (a single target pixel), a Pearson correlation test can be constructed as

$$\psi_4(\underline{T}(\underline{x}_j)) = |\hat{a}_t| = \frac{|\sum_{i=1}^{N-1} T(\underline{x}_j; \rho_i) T^*(\underline{x}_j; \rho_{i+1})|}{\sum_{i=1}^{N-1} |T(\underline{x}_j; \rho_i)|^2} \stackrel{H_1}{<} \beta. \quad (36)$$

The heuristic used here is that the AR coefficients of clutter will be close to 1 due to the Brownian motion characteristics while the target AR coefficients would be less than 1.

Special Case 2: AR vs. White noise process

The statistical model derived in Section 2 for the multiresolution process showed that a resolution increments process using a specific set of resolutions is a white noise process. Using this increments process \underline{T}' , a test between an AR process in \mathcal{X}_t and white noise ($a'_c = 0$) in \mathcal{X}_c can be devised. The increments process for cultural objects

remains an AR process. However, the driving noise for the AR increments process for the target is now correlated. For the sake of simplicity, we will approximate the increments process for the target as AR with an *iid* noise source. In the derivation of this detector, a simple replacement of the original process T by T' is made. The multiresolution increments process can be described as

$$T'(\underline{x}_j; \rho_i) = a'_c T'(\underline{x}_j; \rho_{i-1}) + c'_i \quad \underline{x}_j \in \mathcal{X}_c \quad (37)$$

$$T'(\underline{x}_j; \rho_i) = a'_t T'(\underline{x}_j; \rho_{i-1}) + t'_i \quad \underline{x}_j \in \mathcal{X}_t \quad (38)$$

where c'_i and t'_i are zero mean Gaussian random variables with variances $\sigma_{c'}^2$, and $\sigma_{t'}^2$ respectively. We are testing between

$$\begin{aligned} H_0 : \quad & a'_c = a'_t = 0, \quad \sigma_{c'}^2 = \sigma_{t'}^2 \\ H_1 : \quad & a'_t \neq a'_c = 0, \text{ or } \sigma_{c'}^2 \neq \sigma_{t'}^2 \end{aligned}$$

The GLRT simply becomes an estimate of the AR coefficient in region \mathcal{X}_t as

$$\psi_6(\underline{T}') = |\hat{a}'_t| = \frac{|\sum_{\underline{x}_j \in \mathcal{X}_t} \sum_{i=1}^N T'(\underline{x}_j; \rho_i) T'^*(\underline{x}_j; \rho_{i-1})|}{\sum_{\underline{x}_j \in \mathcal{X}_t} \sum_{i=1}^N |T'(\underline{x}_j; \rho_{i-1})|^2} \stackrel{H_1}{>} \beta \quad (39)$$

The processor can be further simplified when $\mathcal{X}_c = \mathcal{X}_t = \underline{x}_j$. This is a single pixel detector. The GLRT becomes the Pearson correlation test in resolution

$$\psi_3(\underline{T}'(\underline{x}_j)) = \frac{|\sum_{i=1}^{N-1} T'(\underline{x}_j; \rho_i) T'^*(\underline{x}_j; \rho_{i+1})|}{\sum_{i=1}^{N-1} |T'(\underline{x}_j; \rho_i)|^2} \stackrel{H_1}{>} \beta \quad (40)$$

A related heuristic test can be constructed which exploits the notion that the clutter increments process is uncorrelated. However, when a cultural object is present, there will be a nonzero correlation coefficient due to the presence of a mean value. This test is an unnormalized autocorrelation test [18]

$$\psi(\underline{T}'(\underline{x}_j)) = \frac{1}{\hat{\sigma}_c^2} \sum_{i=1}^{N-1} T'(\underline{x}_j; \rho_i) T'^*(\underline{x}_j; \rho_{i+1}) \stackrel{H_1}{>} \beta. \quad (41)$$

3.2 Composite Tests

We will base our detection strategies on the increments process $T'(\underline{x}_0; \underline{\rho})$. The cultural object signal model motivates detectors which exploit the difference in the mean when either clutter or cultural objects are present. We will use the same convention as Section 3.1 where we have a target area \mathcal{X}_t and an assumed clutter area \mathcal{X}_c . For the cases examined here $\mathcal{X}_t = \underline{x}_j$, we have a pixel-by-pixel test. The hypotheses under test will be

$$\begin{aligned} H_0 : \underline{T}'(\underline{x}_j) &\sim \mathcal{N}_c(\underline{\mu}, \underline{\Sigma}_{\underline{x}_j}), \quad \underline{\mu} = \underline{0} \\ H_1 : \underline{T}'(\underline{x}_j) &\sim \mathcal{N}_c(\underline{\mu}, \underline{\Sigma}_{\underline{x}_j}), \quad \underline{\mu} \neq \underline{0} \end{aligned}$$

with $\underline{\Sigma}_{\underline{x}_j} = \gamma \sigma_c^2 I$. We also have available $\underline{T}'(\underline{x}_j)$, $\underline{x}_j \in \mathcal{X}_c$ which are surrounding locations assumed to be clutter.

We have chosen the composite test due to the severe variability of SAR signatures when collection geometry and object condition are not known. This is opposed to choosing a specific object signature and designing a matched filter to it. There does not exist a uniformly most powerful test for the above composite hypothesis. The optimal invariant test with respect to scale and orthogonal transformations is the so-called F test [17] which is equivalent to

$$\psi_2(\underline{T}'(\underline{x}_j)) = \frac{|\underline{\Sigma}_{\underline{x}_j}^{-1/2} \underline{T}'(\underline{x}_j)|^2}{N-1} = \sum_{i=1}^{N-1} \frac{|T'(\underline{x}_j; \rho_i)|^2}{(N-1)\gamma \hat{\sigma}_c^2} \stackrel{H_1}{>} \beta. \quad (42)$$

where $\hat{\sigma}_c^2$ is the estimate of σ_c^2 from surrounding cells assuming no target is present, i.e.,

$$\hat{\sigma}_c^2 = \frac{\sum_{i=1}^{N-1} \sum_{\underline{x}_j \in \mathcal{X}_c} |T'(\underline{x}_j; \rho_i)|^2}{M(N-1)}. \quad (43)$$

The distribution of the test statistic is $\psi_2(\underline{x}_j) \sim F(2(N-1), 2M(N-1))$ under H_0 , and $\psi_2(\underline{x}_j) \sim F(2(N-1), 2M(N-1); \sum_{i=1}^{N-1} (\mu_i^2 / \gamma \sigma_c^2))$ under H_1 . Here $F(m, n)$ is the central F distribution with m and n degrees of freedom, and $F(m, n; \eta)$ is

the noncentral F distribution with the same degrees of freedom and noncentricity parameter η . For the case of large $M(N-1)$, the estimate of $\hat{\sigma}_c^2$ is almost exact. The distributions become a scaled central chi-square under H_0 and a scaled non-central chi-square under H_1 .

A number of other detection strategies are often used to screen data. Our baseline test with which we compare the multiresolution test performance is a pixel-by-pixel F test ($\mathcal{X}_t = \underline{x}_j$) applied to single resolution SAR image data at the finest resolution. This test is

$$\psi_1(\underline{T}(\underline{x}_j)) = \frac{|T(\underline{x}_j; \rho_1)|^2}{\hat{\sigma}_c^2} \stackrel{H_1}{>} \beta \quad (44)$$

This test has been used extensively in initial screening algorithms on SAR data[22]. This test searches for bright signatures in relation to the surrounding clutter. This test is the optimal invariant test with respect to scale and orthogonal transformations. This test can also be generalized to a multipixel test as

$$\psi_7(\underline{T}) = \frac{\sum_{\underline{x}_j \in \mathcal{X}_t} |T(\underline{x}_j; \rho_1)|^2}{\hat{\sigma}_c^2} \stackrel{H_1}{>} \beta. \quad (45)$$

3.3 Generalized Matched Filter

The problem we consider is that of detecting a known target at a given spatial location $\underline{x}_0 \in \mathbb{R}^2$ embedded in zero-mean, complex-valued Gaussian clutter. Using the development of Section 2, we can model the SAR image under the target present hypothesis as:

$$H_1 : T(\underline{x}; \rho) = S(\underline{x} - \underline{x}_0; \rho) + C(\underline{x}; \rho).$$

Here, we assume that C is a stationary complex-valued Gaussian process that models ground clutter, while S is a known coherent target signature. The target absent hypothesis, H_0 , is:

$$H_0 : T(\underline{x}; \rho) = C(\underline{x}; \rho).$$

Note that for simplicity we have ignored receiver noise and the fact that in many cases the target typically occludes the ground clutter.

Single-Resolution Detector. In the *single-resolution case*, ρ_0 , the likelihood ratio test is equivalent to computing the following statistic ψ and comparing it to a threshold [26],[28]:

$$\psi = \int T(\underline{x}; \rho_0) f(\underline{x} - \underline{x}_0) d\underline{x} \stackrel{H_0}{>} \beta. \quad (46)$$

Here, f is a solution to the integral equation

$$\int \Sigma(\underline{x}, \underline{y}) S(\underline{y}; \rho_0) d\underline{y} = f(\underline{x}), \quad (47)$$

and Σ is the covariance function for C . The detection problem is nonsingular for a given signal S if the integral equation (47) has a solution [28]. When C is a white process, the solution to (47) is $f = S^*$, and (46) becomes a matched filter detector.

Multiresolution Detector. Suppose, however, that we are allowed to use the SAR image T at multiple resolutions. The multiresolution likelihood ratio test has a form similar to the single-resolution case:

$$\psi = \int \int T(\underline{x}; \rho) f(\underline{x} - \underline{x}_0; \rho) d\underline{x} d\rho. \quad (48)$$

where now, f is a solution to the equation,

$$\int \Sigma(\underline{x}, \rho; \underline{y}, \rho') S(\underline{y}; \rho') d\underline{y} d\rho' = f(\underline{x}; \rho), \quad (49)$$

and Σ is the covariance of $C(\underline{x}; \rho)$.

Figure 6 shows receiver operating characteristics for the single and multiresolution detectors for the signal s shown in Figure 1. In computing the curves for Figure 6, h was a *sinc* function and we imposed the constraint that the finest available resolution ρ is equal to one fourth of the spacing between adjacent scatterers ($\rho \geq D/4$). One can see that the multiresolution setting admits a significant performance gain over the single-resolution case.

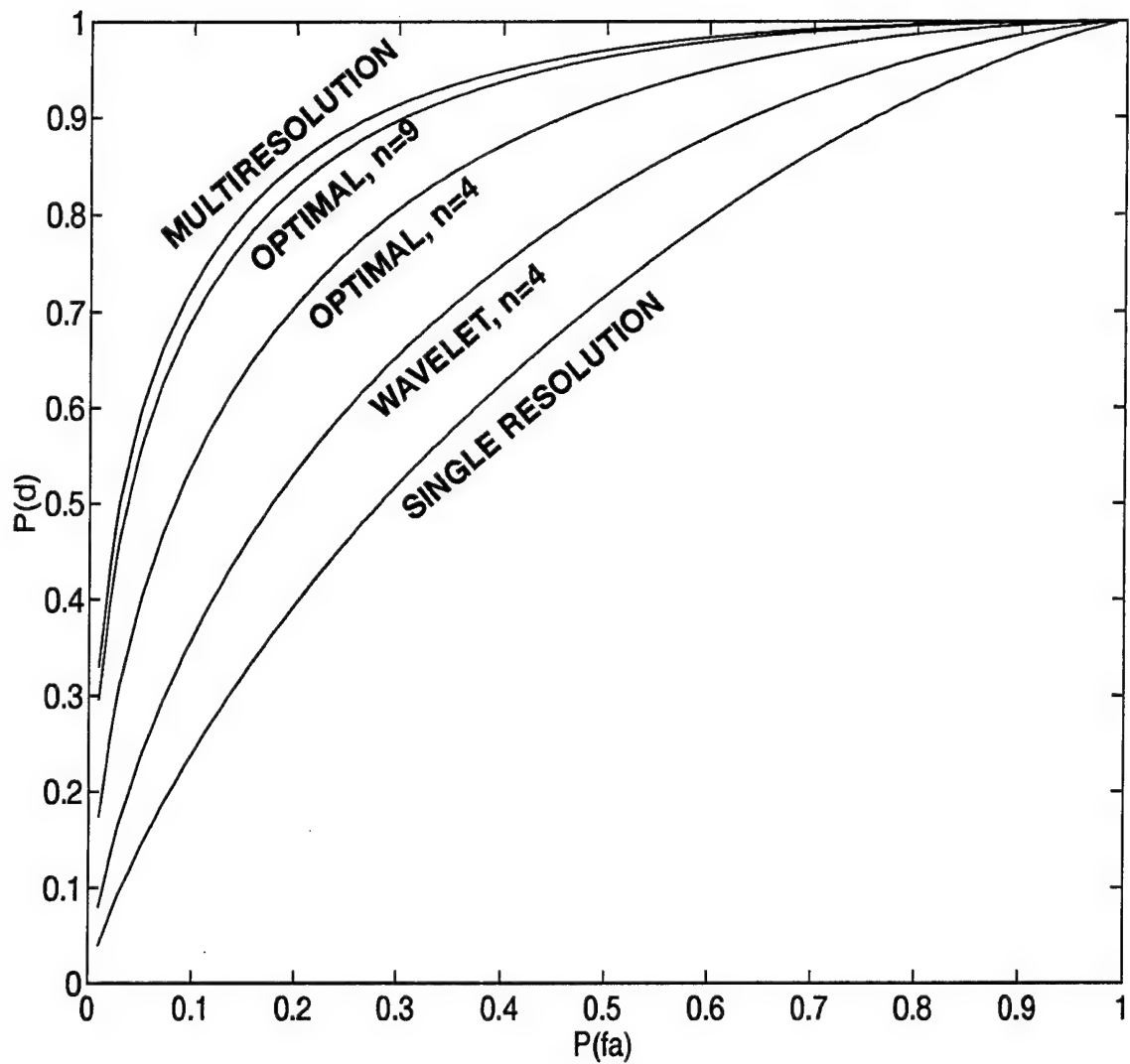


Figure 6: Receiver operating characteristics for the multiresolution detectors using the signal based on three point scatterer target.

Practical Considerations. In comparison to the single resolution detector, Equation (46), the multiresolution detector in Equation (48) requires evaluation of a double integral involving a continuum of resolutions. Practical considerations may lead one to consider using only a *finite* number of resolutions, ρ_1, \dots, ρ_n and replace the statistic T in Equation (48) by some approximating statistic, say ψ_n .

A simple approach to approximating ψ would be to replace the continuously-indexed function $s(\underline{x}; \rho)$ by its (discretized) *wavelet* or *wavelet packet* decomposition [27] and solve the corresponding integral equation for, say $f_n(\underline{x}_i; 2^{-k})$, and use f_n in place of f in (48). From a practical standpoint, wavelets and wavelet packets have the advantage that they are computationally very efficient. However, they are typically limited to dyadic (triadic or similar) resolutions (i.e., integer powers of 2 or 3) and are not necessarily the optimal choices of resolution.

As an alternative, we applied a sampling scheme described in [26] to solve the problem of optimally approximating the test statistic ψ using a finite number of resolutions. This latter sampling scheme requires that one choose resolutions ρ_1, \dots, ρ_n to be the n quantiles of the function $\lambda(\rho) = [\gamma(\rho)\|s(\underline{x}; \rho)\|^2]^{1/3}$, where β is a function depending on Σ , the covariance of $C(\underline{x}; \rho)$, and $\|\cdot\|$ is the L_2 norm. Details of this approach are described in the appendices.

Figure 6 additionally shows the performance of two multiresolution-based detectors that use a finite set of resolutions. The wavelet-based resolution decomposition uses the dyadic resolutions $\{D/4, D/2, D, 2D\}$, while the *optimal multiresolution sampling* scheme for $n = 4$ uses the resolutions $\{0.25D, 0.35D, 0.525D, 1.5D\}$.

Figure 6 shows that the optimal resolution sampling scheme performed much better than the wavelet sampling scheme. Finally, a fifth curve in Figure 6 shows that optimum detector performance is achieved for large enough n .

4 STAR Algorithm Description

4.1 Introduction

This section reviews the algorithm developed under the Strategic Target Algorithm Research (STAR) program executed for Lincoln Laboratories. This algorithm detects extended targets in clutter. It served as a basis for insertion of our multiresolution detection strategy and as a means of performance comparison. Though originally developed for polarimetric data, the ERIM STAR algorithm is general in the sense that it accepts any multi-channel data set such as multiresolution data. We used this algorithm with both single resolution and multiresolution data.

The ERIM STAR detection algorithm exploits statistical knowledge of local clutter. Figure 7 illustrates the algorithm logical structure, which consists of five components: contrast-enhancement/speckle-reduction, CFAR screening, aggregation, feature extraction, and assignment of a cue rating factor (CRF, aka decision statistic). ERIM's detection algorithms assign CRFs to aggregate clusters of pixels (blobs). High CRFs indicate high confidence that a blob is a target-of-interest, while low CRFs indicate low confidence that a blob is a target-of-interest. By definition, the targets-of-interest are the large and medium vehicles; all other man-made and natural objects are uninteresting. Reports of the blob locations, together with their CRFs pass on to scoring software, which evaluates ROC curves by thresholding the CRFs and counting ground-truthed detections and false alarms.

Contrast-enhancement/speckle-reduction (CESR), is a pixel-by-pixel operation that exploits second-order statistical properties of multi-channel clutter to produce an enhanced real-valued image from the complex-valued multi-channel measurements. By simultaneously increasing the contrast between targets and surrounding clutter and minimizing clutter speckle (without compromising resolution), CESR improves the capability of CFAR screening to highlight target pixels and reject false alarm pixels. Empirical observation indicates that the upper (but not lower) tails of the clutter distributions of the CESR image can be conservatively (in terms of false alarm rate)

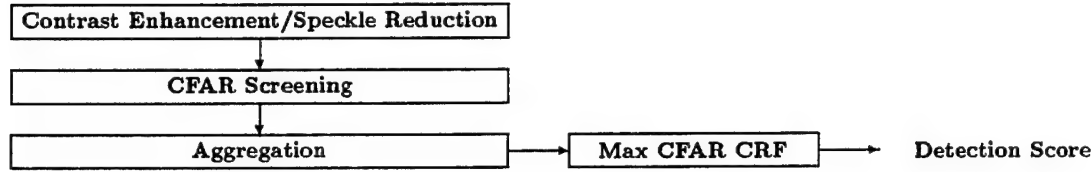


Figure 7: Original STAR multi-channel contrast-based detection algorithm flow.

modelled as log-normal. Thus we evaluate a CFAR statistic which measures the local contrast (in local standard deviations from the local mean) of the log CESR image. We employ a comparatively low screening threshold that, based on log-normal theory, predicts a false alarm rate of 10^3 FA/ km^2 . This threshold provides at least one CFAR hit on 98% of the deployed large and medium training targets.

Obscured targets generally produce multiple disconnected CFAR hits in close proximity to one another. Thus, we employ a cascade of dilation and skeletonization operators to aggregate nearby (on a target-sized scale) CFAR hits into connected, multi-pixel (spatially distributed) blobs. Ideally, the size and shape of target-induced blobs reflect the size and shape of the underlying targets; obscuration generally reduces the fidelity of the blob size and shape. False alarm rates based on counting nearby CFAR hits that are merged into aggregate blobs, each characterized by their maximum CFAR statistic, are an order of magnitude lower, typically, than those based on counting individual-pixel CFAR hits.

The detection algorithm assigns the maximum value of the CFAR statistic over the pixels comprising the blob. Our detection algorithm combines the notion of pixel-by-pixel CFAR with aggregation. The detector exploits only the maximum available contrast against local clutter as its measure of “targetness”.

The discriminator exploits more characteristics, or features, of interesting targets

than does the detector. Consequently, the discriminator does a better job of rejecting culturally-induced false alarms, such as buildings, the small target, uncamouflaged vehicles on roads, trihedrals, sensor artifacts, etc., than the detector. In addition, the discriminator does a better job of rejecting clutter-induced false alarms than the detector.

The multiresolution detection strategies effectively replace the front end of the STAR algorithm which is concerned with the pixel-by-pixel detection operation. This is shown in Figure 8. Our conjecture is that using detection strategies based on multiresolution signatures, enhanced target/clutter contrast will be obtained. As a means of testing that conjecture, the threshold values remained constant when single resolution and multiresolution data was used. Another conjecture was that a higher percentage of pixels will be detected over the extended target using multiresolution data. This conjecture would imply that a redesign of the post pixel-by-pixel detection morphology would be in order. The morphology would be made less aggressive. We decided, however, to leave the morphology unaltered since we found few instances of multiple targets merging together. Lastly, we replaced the discriminant module of the STAR algorithm which is based on a quadratic discriminator by a tree structured classifier. The quadratic discriminator is fundamentally based on the joint Gaussian nature of the feature vector. The tree structured classifier does not make this assumption which we felt was more realistic with respect to the statistics of the derived features.

4.2 Algorithm Description

4.2.1 Contrast Enhancement

Poor contrast between targets and background clutter, together with speckle, are major factors which limit performance of automated SAR image-based target detection algorithms. Contrast-enhancement/speckle-reduction (CESR) techniques exploit the multi-channel data available at each pixel to maximize contrast and minimize speckle

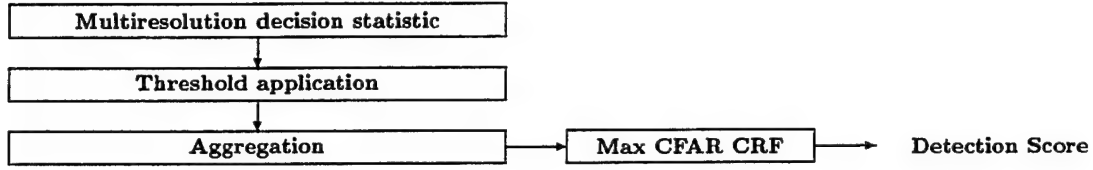


Figure 8: Algorithm flow of multiresolution/STAR detection algorithm.

while preserving resolution. A vector \underline{T} of complex-valued data is associated with each pixel in a multi-resolution SAR image. Conversely, we can also use the multiresolution increments process \underline{T}' . CESR techniques combine these measurements into a single real value that exhibits maximum contrast with respect to background clutter, and which represents an optimum estimate of the intensity of a Gaussian speckle-modulated field. The CESR process evaluates a quadratic form involving the data vector at each pixel and the inverse of the typical clutter covariance matrix. Both contrast-enhancement and speckle-reduction improve performance of the initial CFAR detection stage. Improving the target-to-clutter ratio effectively increases the separation between the image target and clutter means, while reducing speckle decreases the standard deviation of the clutter (and maybe the target). Here we provide the contrast-enhancement motivation for the CESR processor; Novak [1] provides a speckle-reduction motivation.

Consider, at each pixel \underline{x}_j , a linear combination of the data available at that pixel: $\underline{g}^t(\underline{x}_j)\underline{T}(\underline{x}_j)$; $\underline{g}(\underline{x}_j)$ is a vector of spatially-adaptive filter coefficients. The contrast ratio (target-to-local-clutter ratio) at the output of this filter is:

$$C(\underline{x}_j) = \frac{|\underline{g}^t(\underline{x}_j)\underline{T}(\underline{x}_j)|^2}{E_{\underline{x}_i \in \mathcal{X}_c}(|\underline{g}^t(\underline{x}_j)\underline{T}(\underline{x}_i)|^2)}. \quad (50)$$

The $E_{\underline{x}_i \in \mathcal{X}_c}()$ operator denotes expectation over an annular twice-target-sized spatial region \mathcal{X}_c centered on pixel \underline{x}_j . $\Sigma_{\underline{x}_j} = E_{\underline{x}_i \in \mathcal{X}_c}(T(\underline{x}_j)T^t(\underline{x}_j))$ denotes the local clutter covariance matrix. A simple eigenvalue argument establishes the optimum contrast enhancement filter as $\underline{g}(\underline{x}_j) = \Sigma_{\underline{x}_j}^{-1}T(\underline{x}_j)/\|\Sigma_{\underline{x}_j}^{-1/2}T(\underline{x}_j)\|$. The maximized contrast produced by this data-dependent filter is $T^t(\underline{x}_j)\Sigma_{\underline{x}_j}^{-1}T(\underline{x}_j)$. The contrast-enhanced image, in amplitude units, is

$$i_{CE}(\underline{x}_j) = \sqrt{T^t(\underline{x}_j)\Sigma_{\underline{x}_j}^{-1}T(\underline{x}_j)}. \quad (51)$$

Covariance matrices for a wide variety of natural clutter, evaluated over twice-target-sized averaging windows, differ principally in terms of their scaling level. The relative levels of the covariance matrix components vary only slightly with terrain type. Thus, we assume that the local clutter covariance is of the product-model form $\Sigma_{\underline{x}_j} = \sigma^2(\underline{x}_j)\hat{\Sigma}_{\underline{x}_j}$, where $\hat{\Sigma}_{\underline{x}_j}$ is normalized so that its span is one. Thus the maximized local contrast is

$$C_{max}(\underline{x}_j) = \frac{T^t(\underline{x}_j)\hat{\Sigma}_{\underline{x}_j}^{-1}T(\underline{x}_j)}{\sigma^2(\underline{x}_j)}. \quad (52)$$

To preserve image level, we multiply the contrast by the local clutter cross-section $\sigma^2(\underline{x}_j)$. Thus the contrast-enhanced image, in amplitude units, is

$$i_{CE}(\underline{x}_j) = \sqrt{T^t(\underline{x}_j)\hat{\Sigma}_{\underline{x}_j}^{-1}T(\underline{x}_j)}. \quad (53)$$

If we use the covariance matrix for the increments process derived in Section 2 ($\hat{\Sigma}_{\underline{x}_j} = I$), the contrast image simply becomes

$$i_{CE}(\underline{x}_j) = \|T(\underline{x}_j)\| \quad (54)$$

which is the F test described in Section 3.

4.2.2 CFAR Screening

Constant false alarm rate (CFAR) screening is a standard technique for isolating pixels that appear statistically different than the surrounding pixels in an inhomogeneous gray-scale image. In our application, we seek to isolate those pixels that exhibit positive contrast with respect to the local clutter in excess of what one would anticipate, with specified probability, from locally homogeneous clutter. Our detection and discrimination algorithms use these “hits” to trigger the aggregation and feature extraction.

To establish ball-park pixel-by-pixel screening false alarm rates, it is conservative and convenient to model the upper (but not lower) tails of locally homogeneous log CESR clutter as Gaussian. That is, if we measure the local mean and standard deviation of homogeneous log CESR clutter, the associated Gaussian density lies slightly above the upper tail of the clutter histogram. Based on this model, the screener evaluates the two-parameter CFAR statistic $\frac{i(\underline{x}_j) - \mu(\underline{x}_j)}{\sigma(\underline{x}_j)}$. When the Gaussian log-CESR model is accurate, the CFAR statistic is a unit variance, zero mean Gaussian random variable. Thresholding the CFAR statistic (local contrast) at threshold β leads to a per-pixel false alarm probability of $P_{FA} = \frac{1}{2}\text{erfc}(\frac{\beta}{\sqrt{2}})$. The corresponding false alarm rate (FA/km²), assuming M independent pixels per square kilometer, is

$$\text{FAR} = \frac{M}{2}\text{erfc}(\frac{\beta}{\sqrt{2}}). \quad (55)$$

For resampled spotlight and stripmap ADTS data, $M = 16 \times 10^6/\text{km}^2$. Thus, our CFAR screening threshold of 3.84 corresponds to 10^3 FA/km², based on log-normal theory. We also use a CFAR threshold of 2.50 to determine the prominent scatterers on an aggregate blob; this corresponds to 10^5 FA/km².

The CFAR screening algorithm uses a rectangular annular window, centered on the test pixel, to define the clutter surrounding the test pixel, and evaluate its mean and standard deviation. The window is one pixel thick, with a half-width equal to the length of a large target (80 pixels). The window size ensures that pixels on one end

of the target will not corrupt clutter statistics when the test pixel is at the other end of the target. Subject to this constraint, the window incorporates the pixels nearest to the test pixel, which are most representative of local clutter.

4.2.3 Aggregation

The goal of aggregation is to consolidate nearby disjoint hits from CFAR screening into connected blobs which contain the prominent scattering centers together with a minimal number of surrounding pixels. To accomplish this goal, we use two morphological operators: dilation and skeletonization. Dilation grows disjoint regions together, while skeletonization eliminates many of the non-prominent pixels introduced by dilation yet preserves component connectivity. The effect of our aggregation operations is to produce a blob which resembles a hull around the CFAR hits that extends roughly seven pixels around those hits.

Dilation evaluates the local maximum of a gray-scale image over a binary-valued sliding structuring element, which is analogous to the impulse response support of a linear filter. The STAR aggregation algorithm applies a 5×5 octagonal structuring element to the binary-valued hits produced by initial CFAR screening. Each dilation effectively increases the diameter of a CFAR hit region by four pixels. Disjoint CFAR hits separated by less than four pixels grow together into a merged region. We perform a cascade of seven dilations, thereby merging (aggregating) CFAR hits that are separated by less than 28 pixels (35% the length of a large target) into a single blob.

While dilation does merge prominent scatterers into aggregate blobs, the resulting blobs contain many surrounding non-target and/or non-prominent pixels. To eliminate these, one could use an erosion operator, which evaluates the local minimum over a sliding structuring element. Unfortunately, erosion is the inverse of dilation, and could re-fragment the blob. Instead, for binary-valued images, a skeletonization operator exists, which is analogous to erosion, but contains a test to avoid fragmentation. We utilize an 8-way connectivity skeletonization operator, which is analogous

to an erosion that uses a 3×3 square structuring element. We perform a cascade of seven 8-way skeletonizations after the seven dilations.

A cascade of dilations followed by a cascade of skeletonizations is an easy way to merge nearby CFAR hits into aggregate blobs. We chose the number of operators in the cascade, seven, empirically on single resolution data. Had we chosen a very low CFAR screening threshold, it would have been necessary to employ less aggressive aggregation to avoid merging the entire scene, or closely-spaced targets due to the high false alarm rates. Conversely, had we chosen a very high CFAR screening threshold, it might have been necessary to employ more aggressive aggregation to connect more widely-spaced target components.

5 Evaluation of Multiresolution CFAR Performance

This section will outline the empirical evaluation of the various detection strategies using multiresolution signatures. We examine two cases, synthetically generated target and clutter signatures, and the use of actual collected SAR data of targets and terrain through the STAR data set augmented by ERIM DCS data. The simulated target signature is a SAR simulation of a Howitzer target produced by the Synthetic Radar Image Model (SRIM) software. The synthetic background clutter is a homogeneous complex Gaussian random process. The collected data incorporated collected targets in various netted states and deployments, natural homogeneous and inhomogeneous clutter, and cultural clutter. Section 5.1 discusses the methodology by which we generated the synthetic scenes. Section 5.2 provides results and trade studies of the multiresolution detection strategies when applied to synthetic scenes. Section 5.3 explains the results obtained when the multiresolution detection strategies were applied to the collected SAR data sets. Finally, section 5.4 explains the discrimination algorithm that we used in these studies and the results that were obtained.

5.1 Synthetic Scene Generation

The images used for synthetic test data were created by embedding 18 synthetic radar image model (SRIM) generated targets into a simulated homogeneous clutter background. The targets were 32x32 pixels in size at the finest resolution. The targets were multiple realizations of a towed artillery piece. Each realization of the target was generated from different aspect angles which ranged from 0 to 340 degrees in increments of 20 degrees. The clutter was generated as a 512x512 random field. The target-to-clutter ratio was 0dB. The resolution of this simulation was 1 foot. This data set is shown in Figure 9. An additional 512x512 clutter field without embedded targets was also created for false alarm evaluation. The probability distribution of the random field was circular complex gaussian ($p(x_0, y_0) \sim \mathcal{N}_c(0, \sigma_{clut}^2)$) with unit variance ($\sigma_{clut}^2 = 1$). The random process was independent from pixel to pixel at the

finest resolution.

Targets were embedded into the clutter by creating a mask defining the target extent for each realization, then substituting the target pixels for the clutter pixels over the target extent. The target extent was defined by thresholding the target signature at a level 10dB below the average target power. Target pixels which fell below the 10dB threshold were eliminated. Target pixels which were below the threshold but were wholly contained within the target support were allowed, however. This contiguous pixel set defines the support (pixel set) of the target. For our discussion, we will call this pixel set \mathcal{X}_{tgt} . The target pixels contained in the mask were then scaled to unit average power. Areas of the 512x512 homogeneous clutter field conforming to specific target masks were extracted. Target signatures were then inserted into the extracted areas replacing the clutter signatures by the embedded target signatures.

The dynamic range of the target was also constrained to a level which was 10dB above the target's average power defined as

$$\sigma_{tgt}^2 = \frac{1}{M} \sum_{\mathbf{x}_j \in \mathcal{X}_{tgt}} |S(\mathbf{x}_j)|^2 \quad (56)$$

where M are the number of target pixels for that particular target. This was done to reduce large glints in the target signature.

The target-to-clutter ratio (TCR) of the images are determined from the average powers of the target and clutter and is defined as

$$TCR = \frac{\sigma_{tgt}^2}{\sigma_{clut}^2}. \quad (57)$$

To achieve a particular TCR, the target average power will be held constant whereas the clutter will be scaled by σ_{tgt}^2/TCR (recall that the original $\sigma_{clut}^2 = 1$). The targets are then embedded into clutter with the appropriate variance for the chosen TCR. The TCR used in our studies was 0dB.

The mask used to define the extent of the embedded targets are also used to evaluate the output of the detectors. The mask is used to determine whether or not

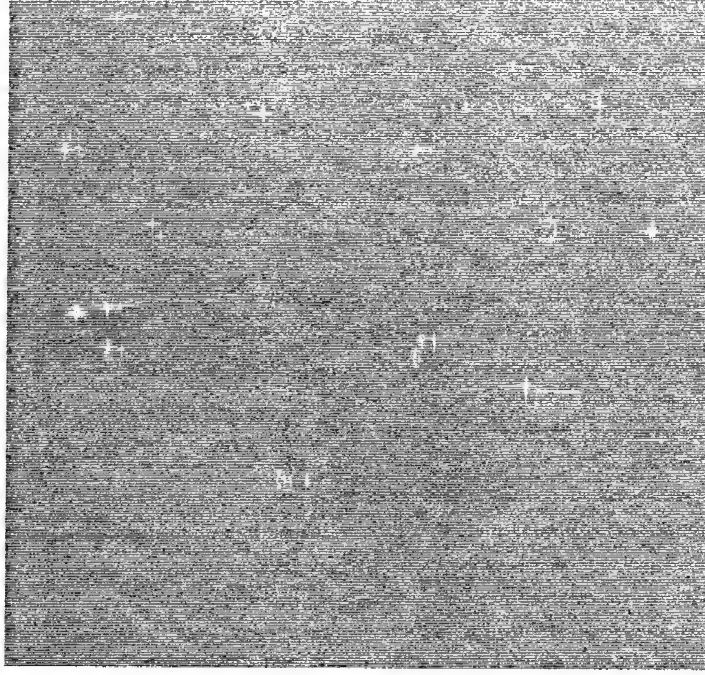


Figure 9: Portion of the data set used in the detector performance studies showing embedded targets with a 0 dB target-to-clutter ratio.

a detected pixel corresponds to a true target pixel or is a clutter false alarm. The probabilities of detection and false alarm are then estimated from the detector output as

$$\hat{P}_D = \frac{\# \text{ of detected pixels} | \text{target}}{\# \text{ of target pixels}} = \frac{n_{tgt}}{N_{tgt}} \quad (58)$$

$$\hat{P}_{FA} = \frac{\# \text{ of detected pixels} | \text{clutter}}{\# \text{ of clutter pixels}} = \frac{n_{fa}}{N_{cl}}. \quad (59)$$

The finite test sample size effects the fidelity of the processor performance estimates. The most traditional method is to look upon the decisions at each pixel as an independent Bernoulli trial [6]. A signal-to-noise ratio ($SNR_{\hat{P}_i}$) of the estimate can be defined as the mean-square to variance ratio and is given by

$$SNR_{\hat{P}_i} \equiv \frac{E^2\{\hat{P}_i\}}{\sigma_{\hat{P}_i}^2}. \quad (60)$$

Description	Notation
Single Pixel/Single Res. F test	$\psi_1(\underline{T}(\underline{x}_j))$
Multi-res. F test(increments)	$\psi_2(\underline{T}'(\underline{x}_j))$
Multi-res. Pearson corr. test (increments)	$\psi_3(\underline{T}'(\underline{x}_j))$
Multi-res. Pearson corr. test	$\psi_4(\underline{T}(\underline{x}_j))$
Multi-res. AR test (AR vs. Brownian motion)	$\psi_5(\underline{T})$
Multi-res. AR test (increments, AR vs. white noise)	$\psi_6(\underline{T}')$
Single Res./Multi-pixel F test	$\psi_7(\underline{T})$

Table 1: List of the various multiresolution detection schemes used in the performance studies.

In our case $N_{tgt} = 1.8 \times 10^4$ and $N_{cl} = 2.5 \times 10^5$. The false alarm estimates will have signal-to-noise ratios of $SNR_{\hat{P}_{FA}} = 2.5 \times 10^3$ for $P_{FA} = 10^{-2}$ and $SNR_{\hat{P}_{FA}} = 24$ for $P_{FA} = 10^{-4}$ which are acceptably large. The signal-to-noise ratios of target detection estimates are computed in a similar manner. $SNR_{\hat{P}_D} \geq 10^5$ for P_D down to 0.1 which is also extremely large.

5.2 Empirical Results – Simulated Data

This section will discuss the results of the various target detection schemes when multiresolution data is used. These detectors were applied to the generated scene discussed in Section 5.1. Table 1 summarises the detectors used in this study. It should be stressed that the Receiver Operating Characteristic (ROC) curves shown in this subsection are based on the *pixel* level detection and not on the extended object. Extended object detection based on collected data is presented in the next subsection.

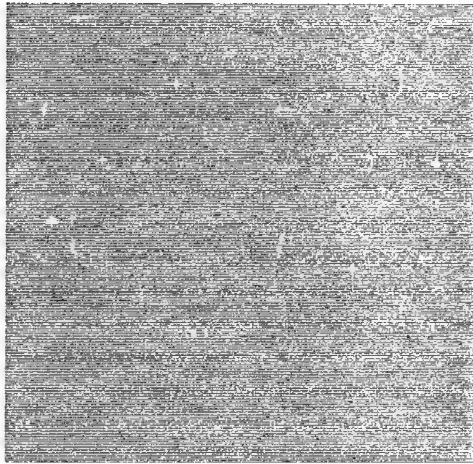
The tests, ψ_1, \dots, ψ_4 used only the signature at the single pixel \underline{x}_j , hence the notation $\psi_n(\underline{T}(\underline{x}_j))$. In the GLRT tests ($\psi_5(\underline{T})$, $\psi_6(\underline{T}')$), \mathcal{X}_t was a 3x3 pixel area. \mathcal{X}_c was a hollow ring of inner diameter 20 pixels and outer diameter 24 pixels. The inner diameter was chosen to be approximately the size of the target. The multi-pixel F test, ψ_7 , also used a 3x3 region for \mathcal{X}_t .

The resolution sampling strategy was chosen to whiten the multiresolution increments process as discussed in Section 2. We used the same resolution sampling strategy for both the original multiresolution process and the multiresolution increments process. The resolution studies only examined azimuth resolution as a running variable. Range resolution was fixed at 1ft. This was due to the conjecture that multiresolution processing may mitigate the requirement for fine resolution collection in azimuth. This has implications on wide area search scenarios since collection time per scene is directly tied to azimuth resolution.

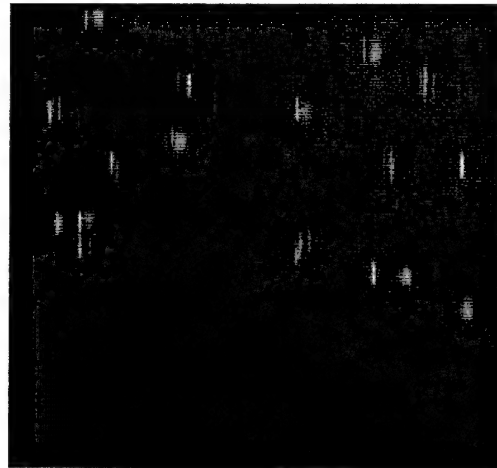
Figure 10 shows the detection statistics of the single pixel detection schemes. In each case, 15 resolutions were used ranging from 1 to 10 ft. The multiresolution tests provide better contrast than the single pixel single resolution test, ψ_1 , which is our baseline. The multiresolution F test, ψ_2 , affords the best contrast between the target set and the clutter. ψ_2 and ψ_3 show a smearing of the signature due to the multiresolution processing. ψ_4 , however, does not exhibit this smearing. Figure 11 shows the empirical Receiver Operating Characteristic (ROC) of three of the single pixel tests: the single resolution/single pixel F test ψ_1 , the multiresolution F test on the increments process ψ_2 , and the Pearson correlation test on the increments process ψ_3 . This figure shows the performance gain provided by the multiresolution signatures. The multiresolution F test, ψ_2 , provided the best result. Performance of the tests increased as the number of resolutions used increases. This is seen by examining the performance of ψ_2 when 6 and 10 resolutions are used. Lastly, note that the multiresolution F test using a 2 ft. starting resolution performed as well as the single resolution F test using 1 ft. data. This has implications regarding the amount of aperture and, hence, search rate, of a SAR system.

The decision statistics showing the increased contrast between the targets and surrounding clutter for the detection schemes using spatial data \underline{T} are shown in Figure 12. Note that the spatial/multiresolution GLRT testing for an AR process vs. Brownian motion, $\psi_5(\underline{T})$, has the best contrast. Figure 13 shows the Receiver Operating Characteristics of the spatial tests ψ_5, ψ_6, ψ_7 and compares them to the baseline ψ_1 .

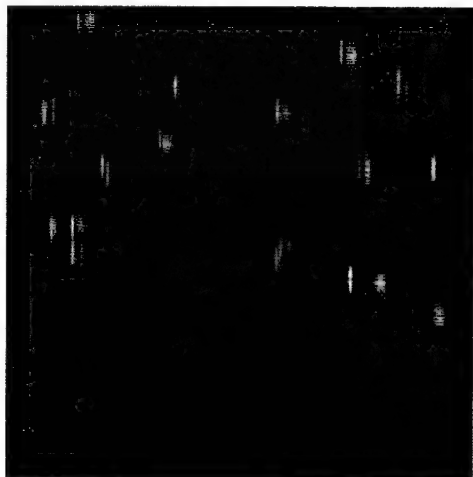
$\psi_1(T(x_j))$ Single Res. 1 ft.



$\psi_2(I'(x_j))$ 1-10 ft., N=15



$\psi_3(I'(x_j))$ 1-10 ft., N=15



$\psi_4(I(x_j))$ 1-10 ft., N=15

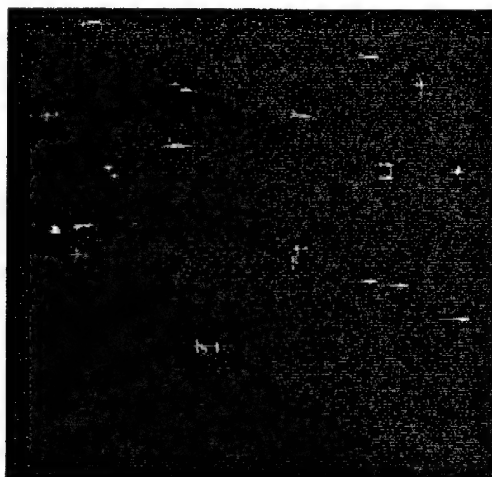


Figure 10: Decision statistics of single pixel tests.

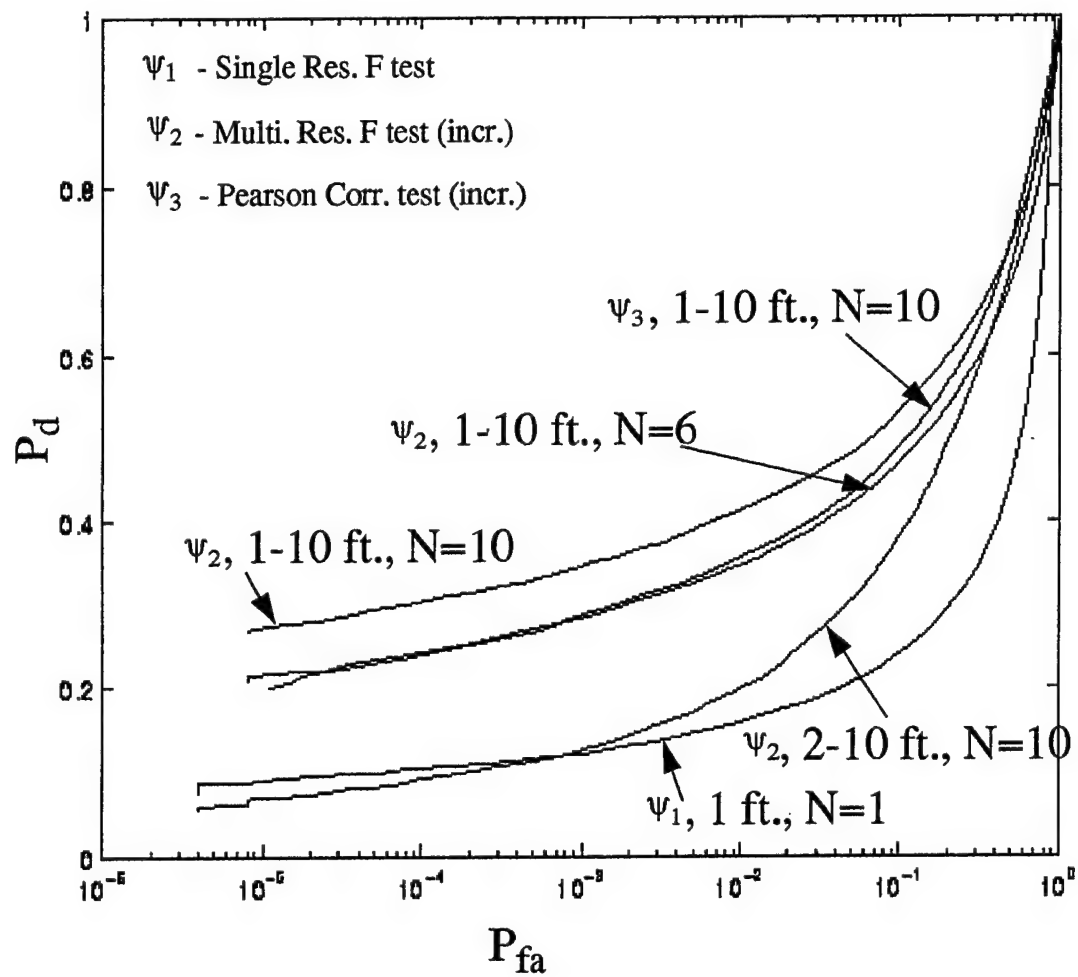


Figure 11: ROC results for various single pixel multiresolution detection schemes.

ψ_5 performed the best and significantly outperformed the single resolution F test, ψ_1 , and the spatial F test ψ_7 . The spatial F test and the multiresolution increments test ψ_6 performed equivalently. The reduced performance by ψ_6 is due to the assumed *iid* driving noise process in the target process model. We expect that if that test is reformulated such that the appropriate driving noise process is accounted for that its performance will be at the level of or better than ψ_5 . Histograms of the various decision statistics are shown in Figure 14. These histograms show the separation increase between the target and clutter signatures when multi-resolution processing is applied.

The next set of figures will explore some of the specific tests more fully and bring out some salient features of multiresolution processing. Figures 15 and 16 show the performance gain as the number of resolutions used in the F test, ψ_2 , increases. Using a starting resolution of 1 ft., there is a significant gain in performance going from 5 to 10 resolutions. The performance gain is smaller when 15 resolutions are used. At 15 resolutions, the performance gain is effectively saturated. Starting from 1.5 ft. resolution, the amount of performance gain is severely reduced. The performance is effectively saturated using 10 resolutions. Figures 17 and 18 show similar results for the P test.

Figures 19 and 20 show the performance degradation of the P test and F test as the starting resolution is degraded. For both the F and P tests, starting resolutions up to 1.5 ft. performed as well as or better than the single resolution F test baseline detector. At coarser starting resolutions, the multiresolution tests outperform the single resolution test at moderate/high false alarm rates.

Figures 21 through 24 show the performance of the tests when the starting resolutions were varied. The number of resolutions were held constant at 15. In each case, the performance of the tests degrade as the starting resolution is coarsened. At a starting resolution of 1.5 ft., the multiresolution F test, ψ_2 , and the Pearson correlation test, ψ_3 , performed the best and were essentially equivalent. They exhibited a significant performance gain over a single resolution F test. At 2 ft. starting reso-



Figure 12: Detection statistics for the various spatial multiresolution strategies.

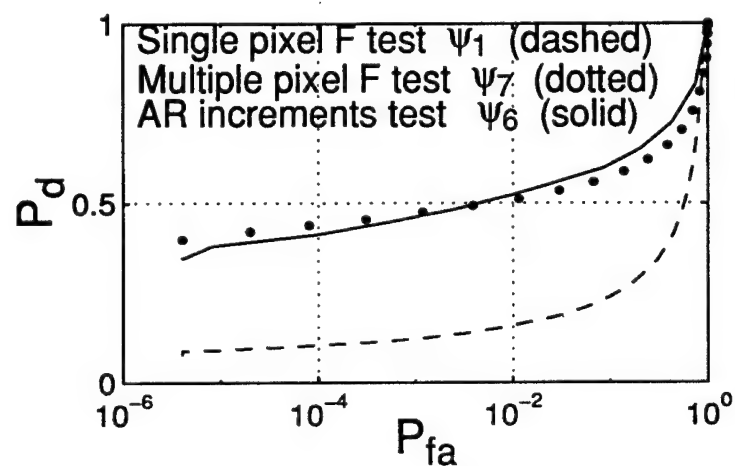
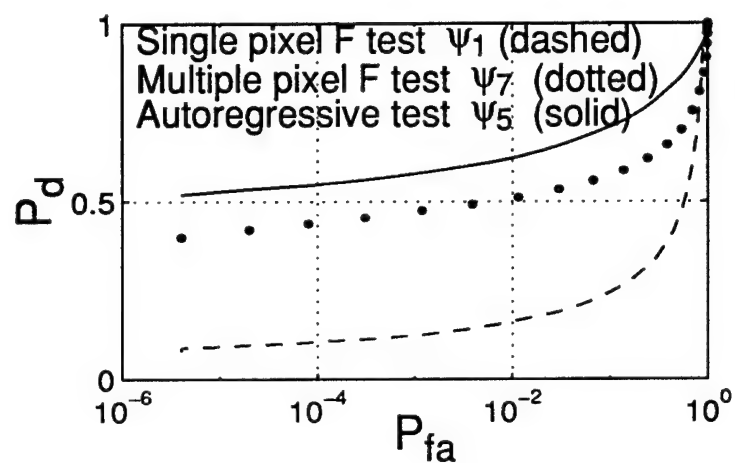
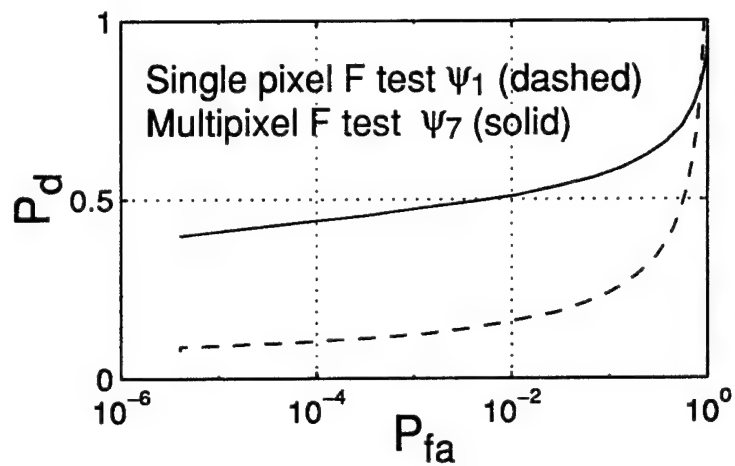


Figure 13: ROC performance of the various spatial multiresolution strategies.

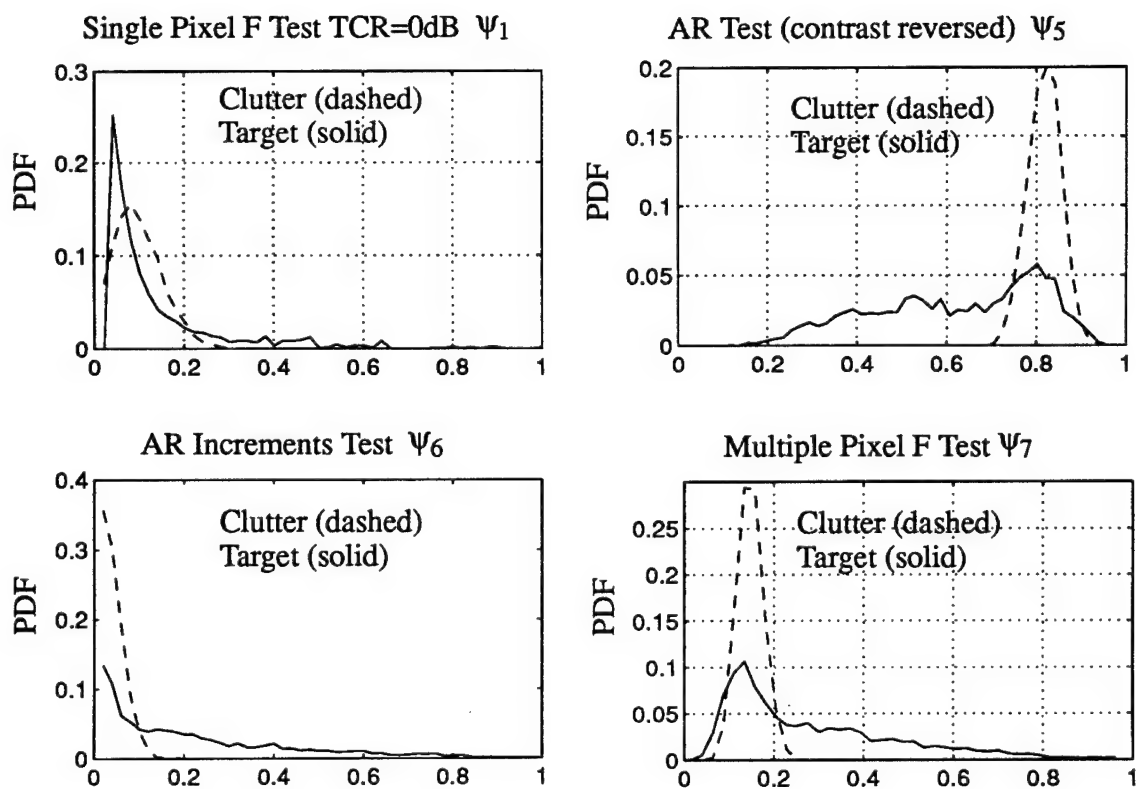


Figure 14: Histograms of the decision statistics for the various strategies.

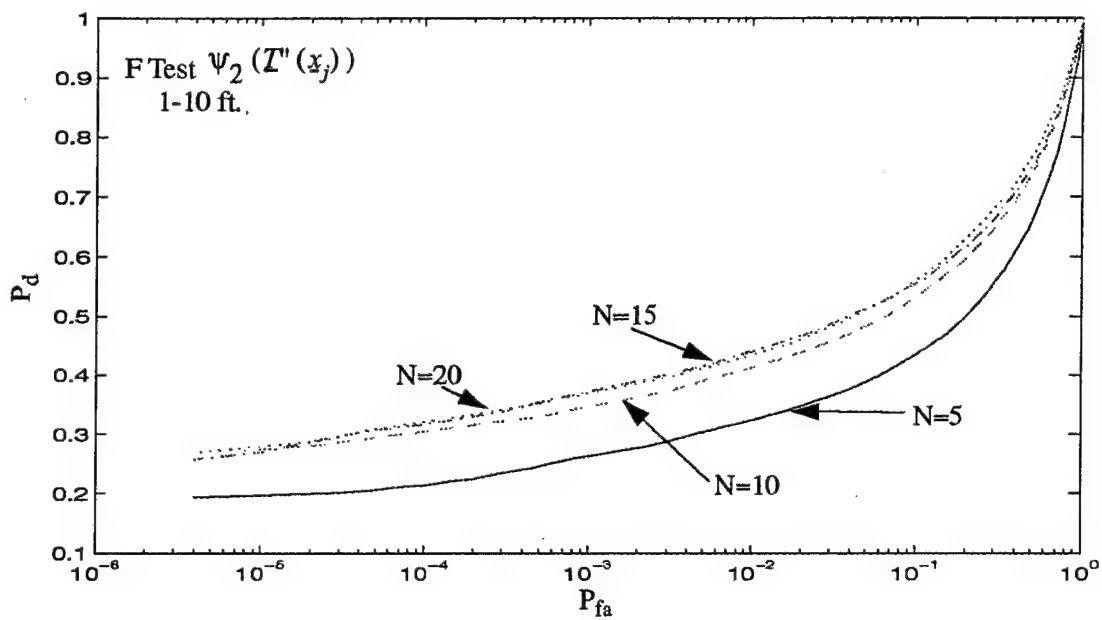


Figure 15: ROC performance of the multiresolution F test as the number of resolutions are increased. Starting resolution is 1ft.

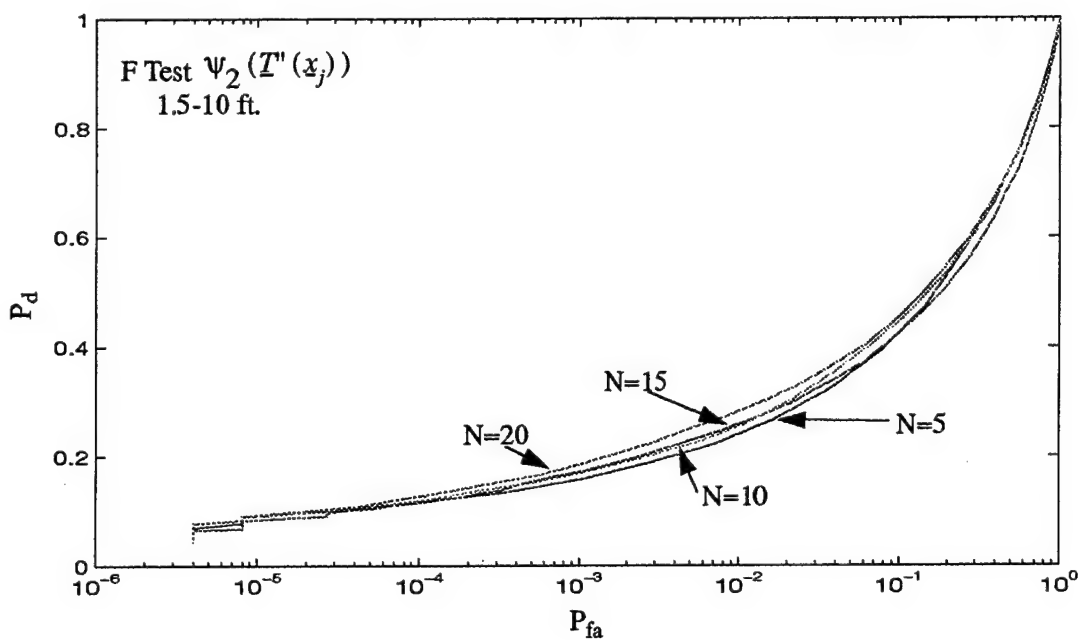


Figure 16: ROC performance of the multiresolution F test as the number of resolutions are increased. Starting resolution is 1.5ft.

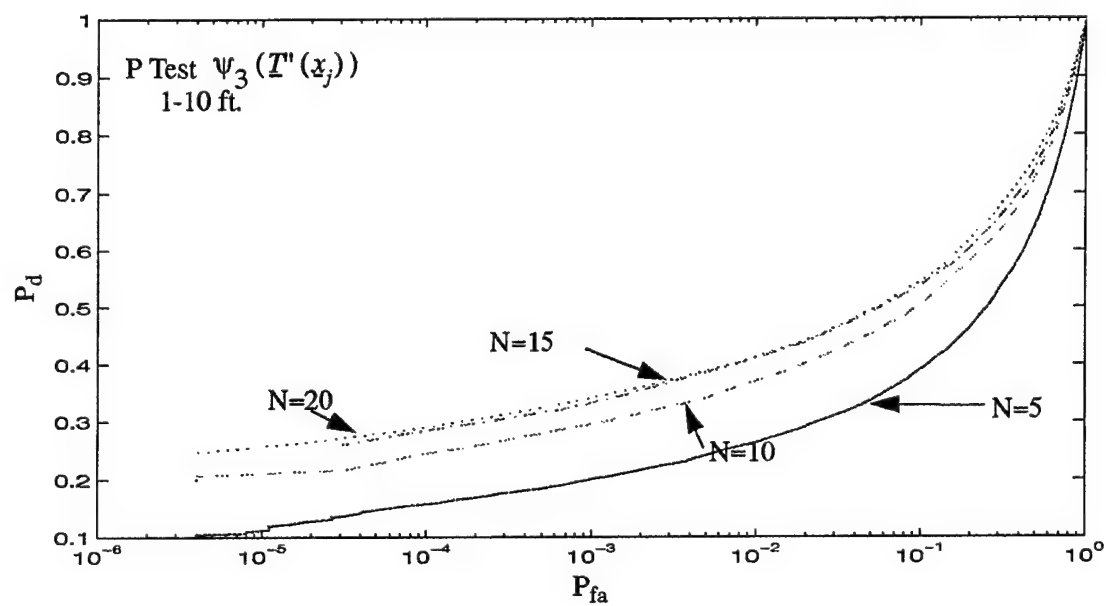


Figure 17: ROC performance of the multiresolution P test as the number of resolutions are increased. Starting resolution is 1ft.

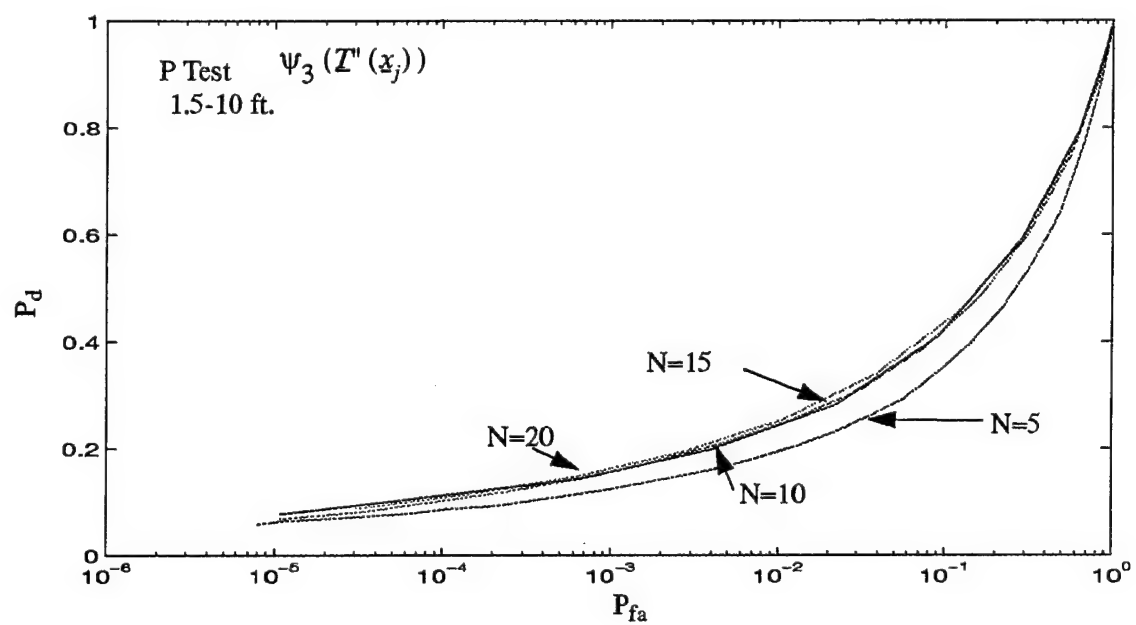


Figure 18: ROC performance of the multiresolution P test as the number of resolutions are increased. Starting resolution is 1.5 ft.

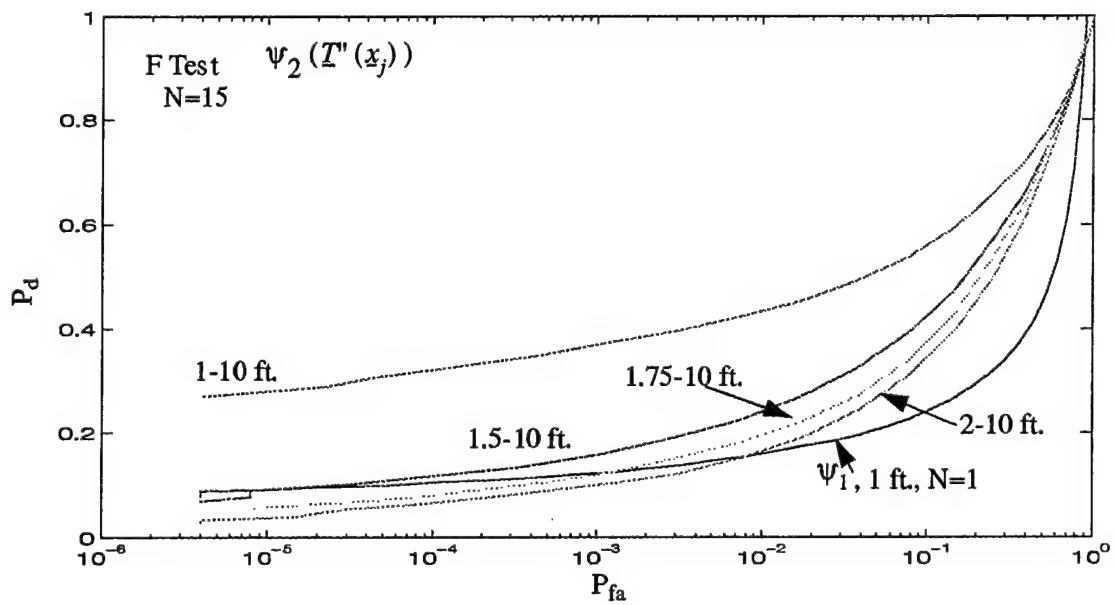


Figure 19: ROC performance of the multiresolution F test as the starting resolution is coarsened.

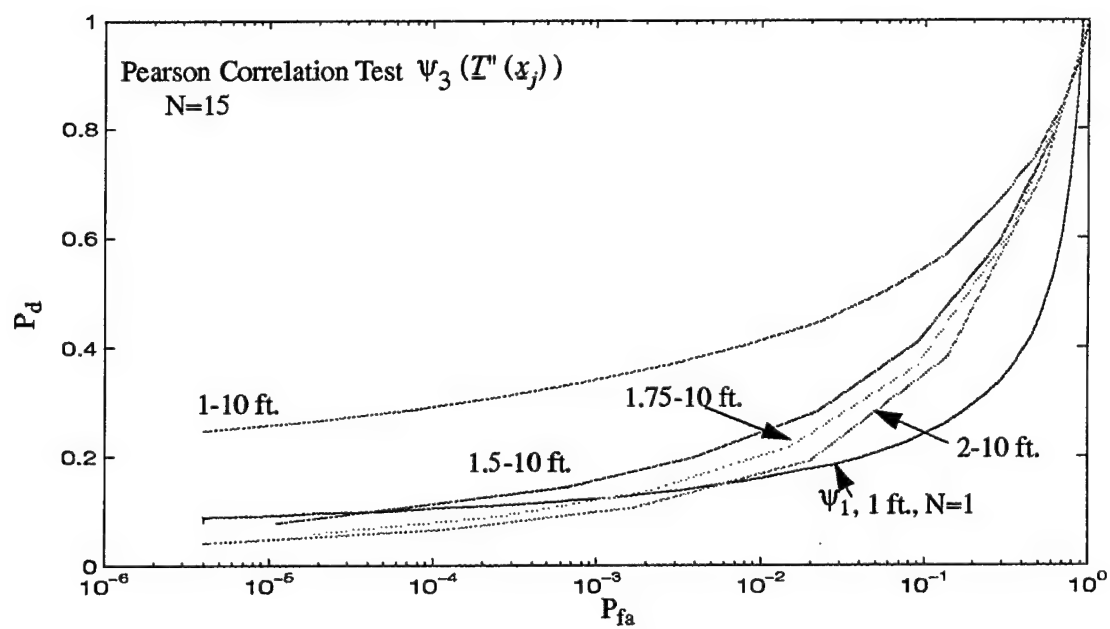


Figure 20: ROC performance of the multiresolution P test as the starting resolution is coarsened.

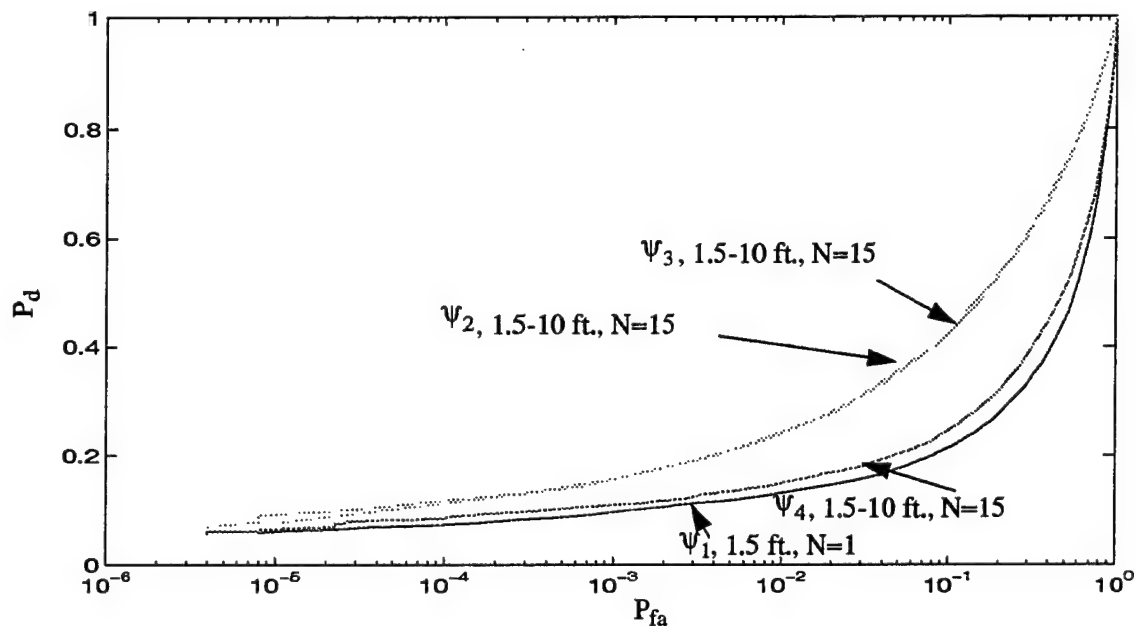


Figure 21: ROC performance of the multiresolution tests for a starting resolution of 1.5 ft.

lution, multiresolution gain is only exhibited at moderate/high false alarm rates. At a starting resolution of 3 ft., there was practically no performance gain provided by the multiresolution tests. Lastly, at a starting resolution of 5 ft., the multiresolution tests performed worse than the single resolution test. The degraded performance at the coarse starting resolutions is to be expected since the law of large numbers will become applicable to the target signatures in this resolution regime.

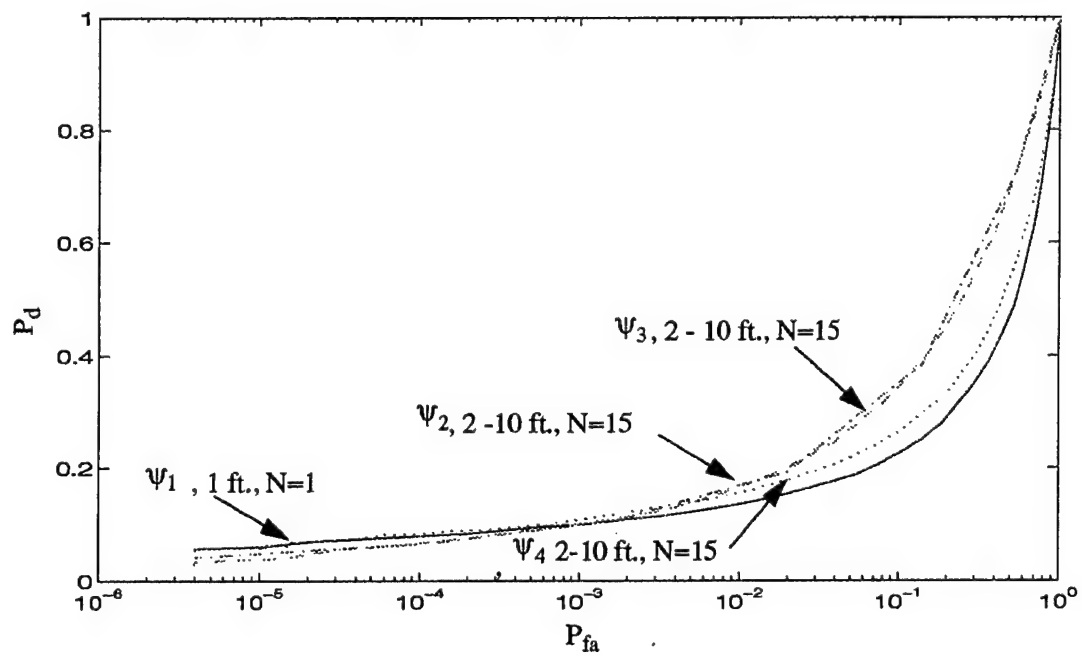


Figure 22: ROC performance of the multiresolution tests for a starting resolution of 2 ft.

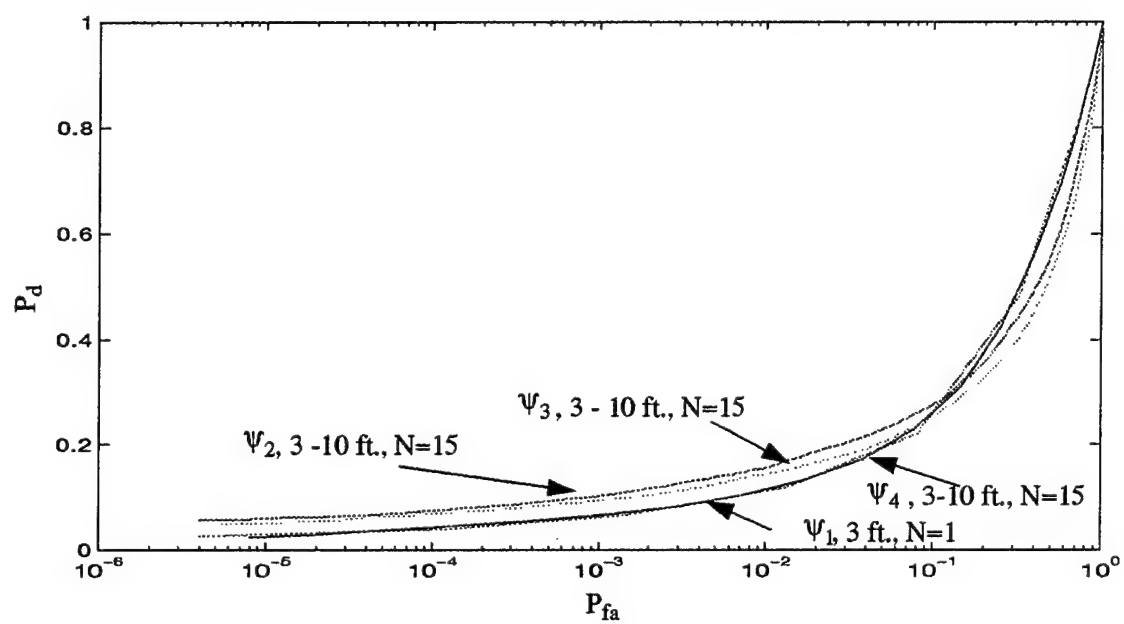


Figure 23: ROC performance of the multiresolution tests for a starting resolution of 3 ft.

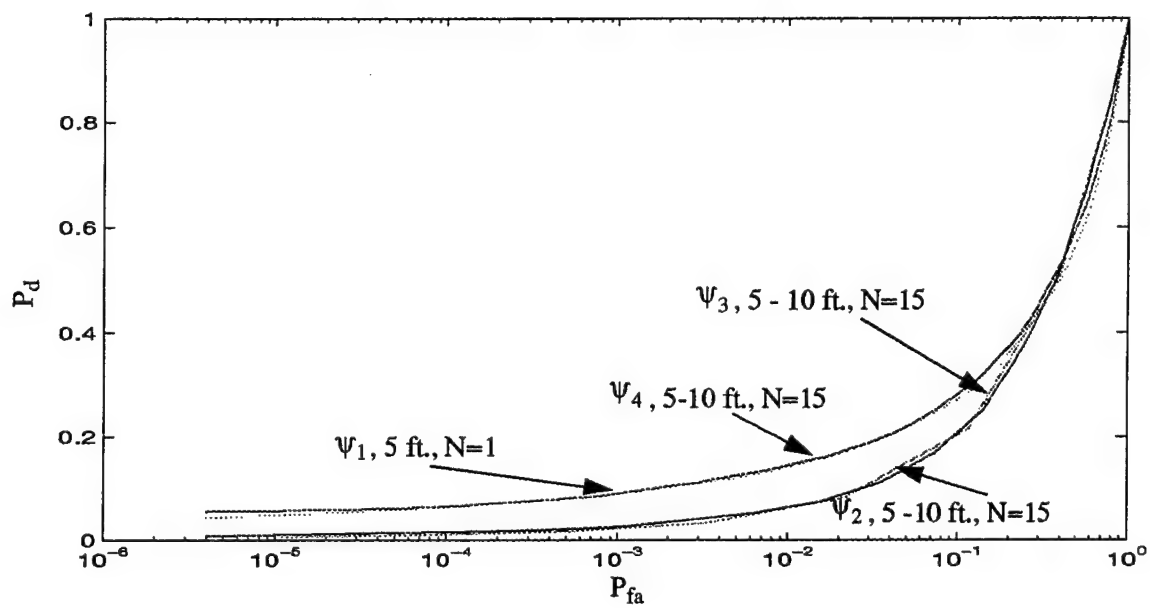


Figure 24: ROC performance of the multiresolution tests for a starting resolution of 5 ft.

5.3 Empirical Results – STAR Data

In this subsection, we describe the results obtained from applying the various multiresolution detection schemes on collected data. The STAR algorithm described in Section 4 is brought to bear on this problem. The STAR data set was also enhanced using ERIM DCS data from Grayling, MI. These data sets were ground truthed with respect to the type and position of the targets within each scene. The scoring that we will show in this subsection is based on the detection of the *extended* targets and not the pixel-by-pixel scoring that occurred in the previous subsection. All non target detections were counted as false alarms regardless of their cause (e.g. placed trihedrals, other cultural objects, etc.). No spatial filtering was performed in the detection stage to eliminate small false alarms. The peak value of the decision statistic within a detection blob was used as the cue rating factor to generate the empirical ROC curves.

The data sets used in these studies were provided by the Lincoln Laboratory ADTS system and the ERIM DCS system. Both systems are fine resolution SAR systems. The ADTS data is the standard data set that was used in the Strategic Target Algorithm Research (STAR) program. This data set was used so we could compare our results with previous detector developments. This data set contained numerous military vehicles in various deployments and netting conditions. There were approximately 780 target realizations in the ADTS data set. This set was augmented by the DCS collections in Grayling, Aberdeen and Eglin. An additional 260 target realizations were made available. The clutter data that was used was from the ADTS sensor. This clutter included both natural and cultural clutter. Approximately 750 square km of clutter was used in the studies.

The TCR gain provided by multiresolution processing on the data sets is shown in Figure 25. This figure shows a rank ordering of the TCR gain between the single resolution F test and the multiresolution F test for both the ADTS and DCS data sets. We define the TCR gain through a general statistical distance metric called the

Hellinger distance. For the two hypotheses

$$H_0 : \quad \text{Clutter } \psi \sim p_0(\psi)$$

$$H_1 : \quad \text{Target } \psi \sim p_1(\psi)$$

the distance (TCR) between the target and clutter distributions is defined as

$$TCR = H(p_0(\psi), p_1(\psi)) = \frac{1}{\sqrt{2}} \left[\int (p_0^{1/2}(\psi) - p_1^{1/2}(\psi))^2 d\psi \right]^{1/2}. \quad (61)$$

For two univariate Gaussian distributions where $p_0(\psi) \sim \mathcal{N}(\mu_c, \sigma_c)$, $p_1(\psi) \sim \mathcal{N}(\mu_t, \sigma_t)$, the Hellinger distance simplifies to

$$H(p_0(\psi), p_1(\psi)) = \frac{1}{\sqrt{2}} \left[1 - \sqrt{\frac{2(1 + \gamma^2)}{\gamma}} \exp \left\{ \frac{(\mu_t - \mu_c)^2}{2\sigma^2(\gamma^2 + 1)} \right\} \right]^{1/2} \quad (62)$$

where $\gamma = \sigma_t^2/\sigma_c^2$. Other metrics such as Mahalanobis distance assume that the mean value or variance are equal between the two hypotheses.

We calculate the Hellinger distance (TCR) between pairs of target and clutter areas of the decision statistics of the detection schemes. For the single resolution case we call this H_{sres} . The multiresolution case is called H_{mres} . We compare the Hellinger distance on the same clutter/target areas to compute gain. The clutter areas chosen were natural clutter areas. Each target realization had a distinct clutter area associated with it. The Hellinger distance is an ‘‘amplitude’’ measure of distance. Therefore, the gain in dB is computed as

$$Gain (dB) = 20 \log H_{mres} - 20 \log H_{sres}. \quad (63)$$

We found an average of 4.2 dB of TCR gain due to multiresolution processing for the ADTS data and an average gain of 3.7 dB for the DCS data.

Figure 26 and 27 shows the ROC performance of the various tests on the STAR data set. Figure 26 compares the performance of various multiresolution detectors

and compares it to the single resolution STAR baseline algorithm ψ_1 . ψ_4 performed the best of all the tests having both the smallest false alarm rate and the largest detection rate. ψ_3 was next followed by ψ_2 . The baseline single resolution test, ψ_1 , performed the worst both in terms of detection rate and false alarm rate. In all cases the starting resolution was 1ft. 15 resolutions between 1 and 10 ft. were used in the multiresolution tests. The detection performance of the single resolution detector saturated at $P_d \simeq .55$. This is due to the setting of the pixel-by-pixel threshold. No pixels were detected on some dim targets. This threshold remained constant when the multiresolution process was used. The multiresolution detectors saturated at $P_d \simeq .78$. The enhanced detection performance demonstrates the TCR gain afforded by the multiresolution processing. The same conclusions can be derived from the DCS data set and is shown in Figure 27.

Figure 28 shows the performance of ψ_1 and ψ_4 as a function of starting resolution. As resolution is coarsened from 1 to 3 ft. the single resolution detector ψ_1 has degraded detection performance. The pixel-by-pixel threshold was held constant. Therefore, as resolution is coarsened there is less peak target energy. For all three resolutions considered in this study, the multiresolution detector significantly outperforms the single resolution detector. The multiresolution detector provides better detectability at 3 ft. than the single resolution detector at 1 ft.

5.4 Results - Discrimination

This subsection describes the results of the discrimination stage of the algorithm. The results obtained in this section are based on the STAR collected data set. Our discrimination algorithm is a false alarm rejection algorithm based on spatial features derived from the binary detection maps that the STAR detection algorithm provides.

5.4.1 Feature Extraction

Here we discuss the rationale and computation of the features extracted for each aggregate blob. Bear in mind that these features are selected based on their potential

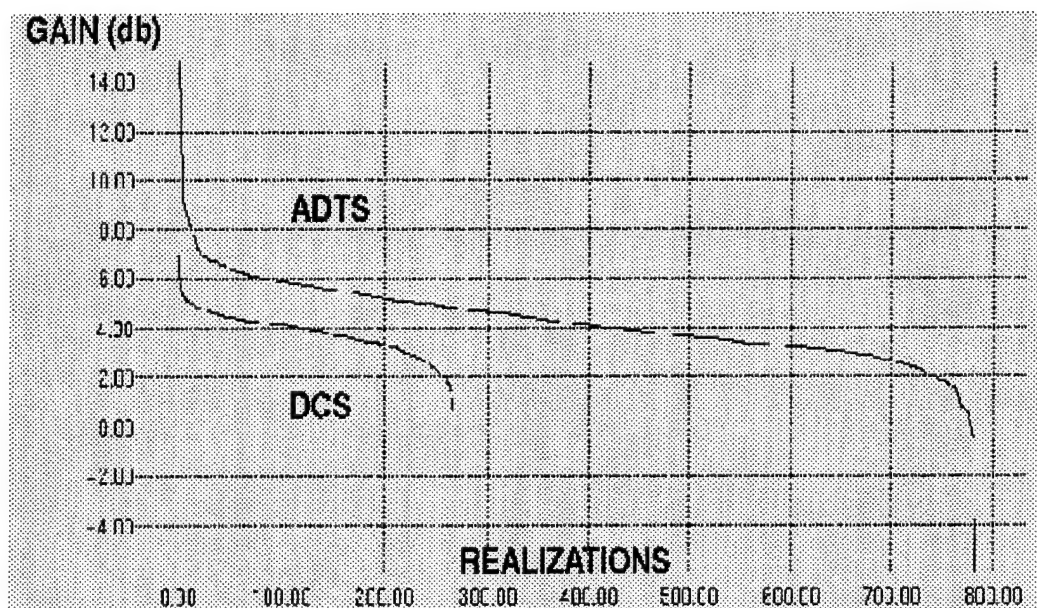


Figure 25: Target to clutter ratio gain afforded by multiresolution processing.

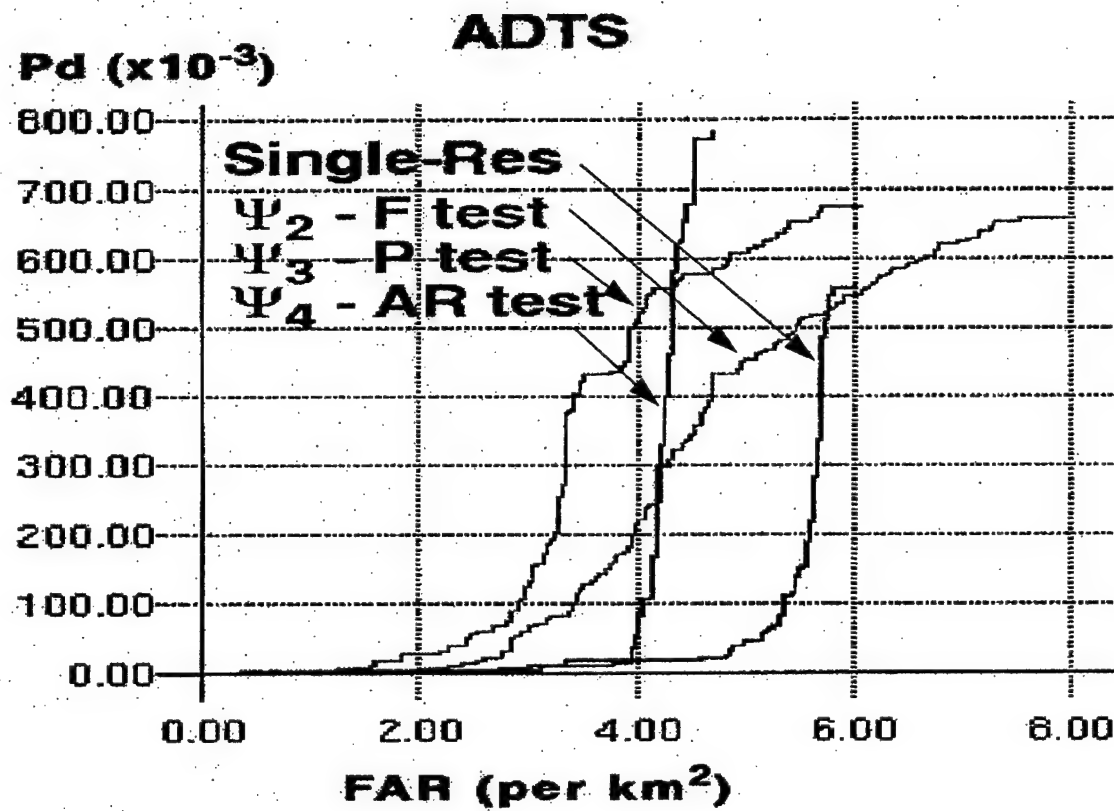


Figure 26: ROC performance of the multiresolution tests applied to the STAR data set.

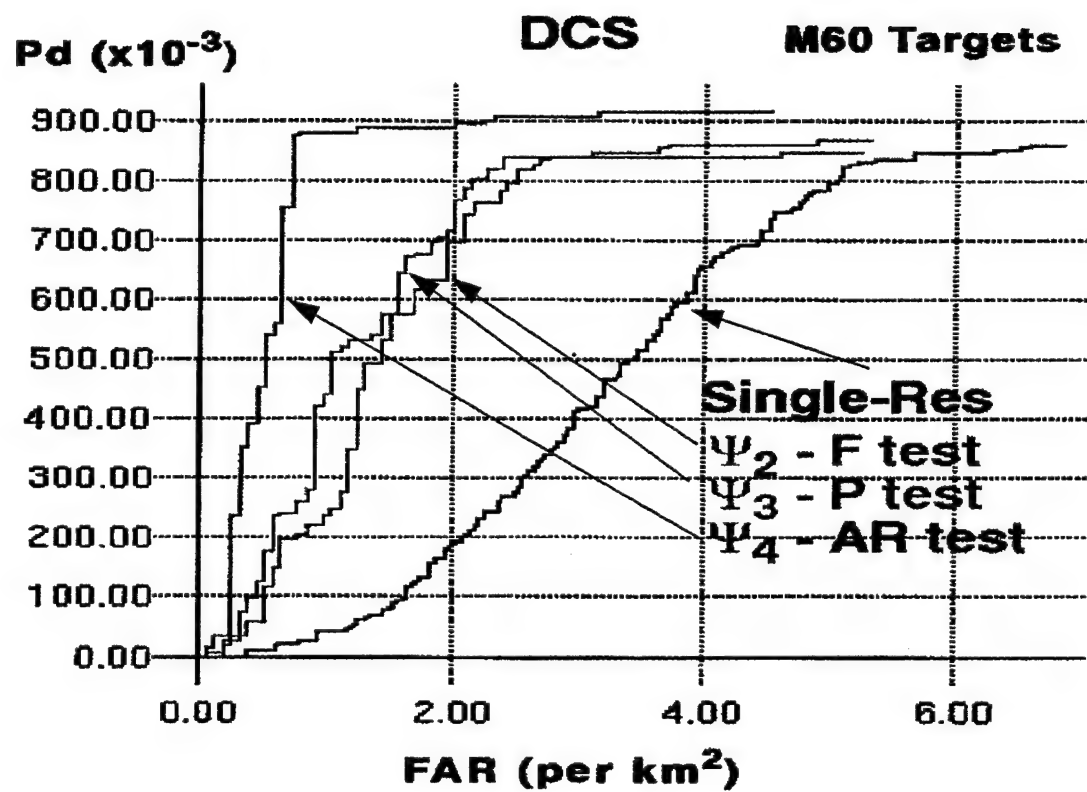


Figure 27: ROC performance of the multiresolution tests applied to the DCS data set.

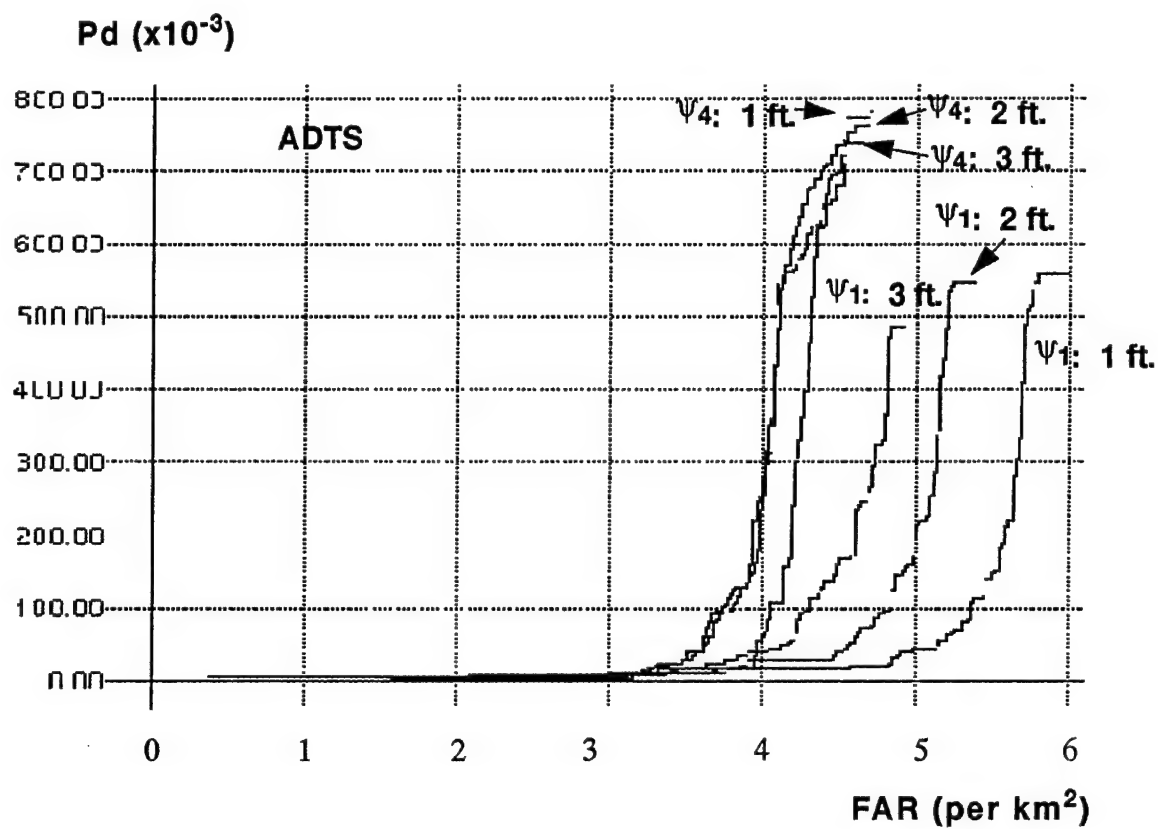


Figure 28: ROC performance of the single resolution and multiresolution AR tests when the starting resolution is coarsened.

to cluster differently on targets-of-interest than on other objects. Table 2 summarizes the features extracted.

Type	Feature
spatial	mass
	diameter
	square-norm. rotational inertia

Table 2: List of features extracted for each aggregate blob.

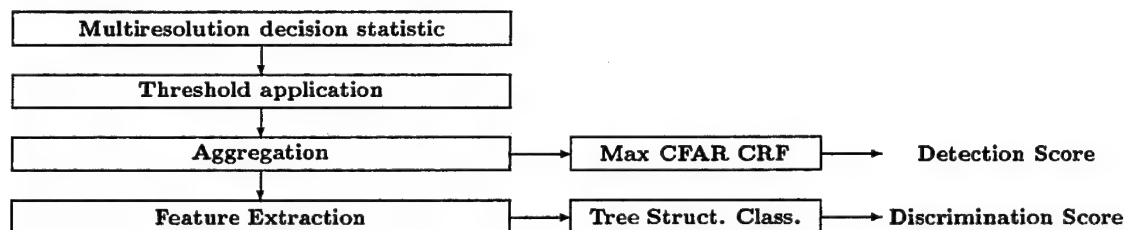


Figure 29: Detection/discrimination algorithm flow of multiresolution/STAR algorithm.

The three spatial features provide distinct measures of coarse spatial properties of an aggregate blob. Mass is the number of pixels in the blob. Diameter measures, roughly, the maximum linear dimension of the blob, in pixels. The diameter feature actually is the integer part of the length of the diagonal of a horizontally or vertically oriented rectangle which just encloses a blob, rounded to the nearest integer. For example, the diameter feature for a 5×7 pixel rectangle, oriented horizontally, is 8. Square-normalized rotational inertia (SNRI) is the second moment of the blob pixel coordinates about the blob center of mass (its inertia), divided by the inertia of an

equal-mass square blob. The expression for SNRI is:

$$\text{SNRI} = \frac{6}{M(M-1)} \sum_{m=0}^{M-1} ((x_m - \bar{x})^2 + (y_m - \bar{y})^2),$$

where the (x_m, y_m) are the coordinates of the m^{th} pixel on the blob, and (\bar{x}, \bar{y}) are the coordinates of the blob center of mass, and M is the blob mass. The SNRI for a horizontally-oriented $L \times W$ pixel rectangle is $\frac{1}{2} \frac{L^2+W^2-2}{LW-1}$. As the length/width ratio (aspect-ratio) of the rectangle increases, the SNRI approaches half the aspect-ratio. In general, SNRI increases as higher proportions of blob pixels become distant from the center of mass. Filled circular blobs exhibit low SNRI (less than one), while long, thin, or annular blobs exhibit high SNRI. Squares exhibit unit SNRI.

5.4.2 Discrimination Algorithm/Results

The discrimination algorithm that we employed in these studies was a feature based tree structured classifier. We only examined binary spatial target features since we conjectured and found that the multiresolution process provides a higher per pixel detection rate. We found that the extended target had more pixel hits on it using multiresolution data than for the single resolution case. We therefore conjectured that the binary spatial features would be a powerful and simple discriminant feature set. We took the tree structured approach due to its nonparametric nature. There are no compelling reasons known to the authors that would lead to any specific parametric model for the three spatial features cited; especially a multivariate Gaussian model.

Tree structured classification approaches are nonparametric in nature. By this we mean that no prior statistical model is assumed for the classes of interest. Rather the processor is trained from collected data. Tree structured approaches have one great advantage over other nonparametric algorithms, however. For large training sets tree structured approaches have been shown to converge to the optimum performance (minimum probability of error) produced by the Bayes classifier [29]. This states that the tree structured classifier would attain the performance of a parametric

processor where the underlying models were completely specified and correct. Many nonparametric algorithms do not exhibit this characteristic since they must assume *a priori* the order of the processor and hope that the data will conform to this order.

A tree structured algorithm is a sequence of binary decisions on data to extract information comprehensively and rapidly. A decision tree is shown in Figure 30. The tree is constructed by repeated splits of the feature space \mathcal{X} into subsets. By feature space, we mean the space of all possible measurement vectors (a measurement vector consists of an ordered group of observables e.g. the vector of binary spatial features of an extended target detection). As shown in the figure, each split is binary (the feature space is split into two subsets). The decision point where a split occurs is called a node.

Trees are normally grown via a steepest descent type algorithm. These algorithms are formally equivalent to the K-means algorithm [30]. These algorithms are iterative and try to optimally split the feature space into two distinct areas at each node. A minimum missclassification error criteria is used as the measure of performance at each node. An initial set of partitions are selected. The missclassification rate is then estimated. The partitions are perturbed until the minimum missclassification rate is found at that node. This procedure is stepwise optimal [31] (no other procedure can do better with respect to minimum error rate at this node).

The decision of when to declare a node terminal is based on a "purity measure." Each node splits the feature space making the resultant data "more pure." Numerous purity measures exist. One of the most often used is [29] is the entropy of the data at that point. If the purity of the data does not change by a predetermined amount that node is defined as a terminal node.

Class labels are easily affixed to terminal nodes. The training and tree growing procedures are supervised. Therefore each input data vector has a label for each of the classes of interest. A simple voting routine where the class with the largest number of training data residing in that partition defines the label of the terminal node.

Figure 30 depicts the results obtained by running the tree structured discrimination algorithm. As shown, for the single resolution case, only 173 of a possible 297 targets were given to the tree structured classifier. This is due to the lower detection rate of the single fine resolution detector ψ_1 . In addition, 365 false alarms were presented to the detector from the single resolution data. Conversely, the AR test ψ_3 had a much higher detection rate so that 261 of the 297 targets were presented to the tree structured classifier. Its lower false alarm rate provided 287 false alarms. A classification tree was grown and optimized for the single resolution and multiresolution cases. The results show that for the single resolution case, 20 false alarms were still classified as target while 4 targets were rejected as false alarms. For the multiresolution case, 14 false alarms were classified as target and 0 targets were rejected as false alarms. These results lend credence that the multiresolution detector provides more target "fill" on extended targets (higher per pixel detection rate) which provides a better discrimination for spatial the features used here. This extra "fill" can be exploited more fully in a discrimination algorithm than that found in the single resolution case.

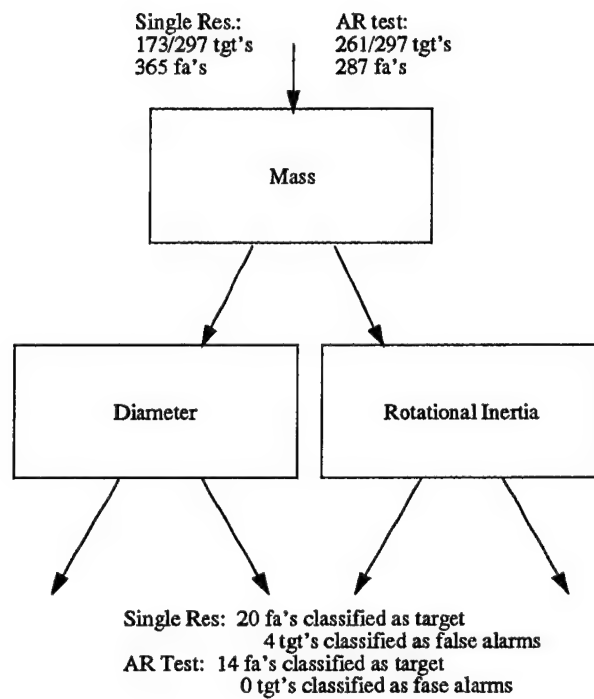


Figure 30: Results of the discrimination algorithm on both the single resolution and multiresolution data.

6 Conclusions

In this section we briefly summarize the main points of the current research. We have attempted to explore multiresolution processing as a means of providing high performance target screening in SAR data. We are exploiting the interference between prominent scatterers in a resolution cell as our discriminant. The hope is that the interference will provide a characteristic signature change as resolution is changed. We formulated statistical models for both clutter and target multiresolution signatures. We showed that the multiresolution clutter process was a Brownian motion process. Using the property of independent increments for Brownian motion processes, we derived a simple resolution sampling strategy that whitens the clutter process. Based on these models, we developed a number of multiresolution tests. These tests consisted of a Generalized Likelihood Ratio approach, composite hypothesis tests and a generalized matched filter. We discounted the matched filter test since it presupposed a specific target signature which we didn't feel conformed to the spirit of first stage screening.

Examining synthetic data, we found that the multiresolution detector far outperforms single resolution detectors on a per/pixel basis. A Generalized Likelihood Ratio approach using a local target area provided the best result. We noted that the performance of the multiresolution detectors saturated at approximately 15 resolutions. We also noted that when the starting resolution was coarsened, the performance of the multiresolution detector suffered. Performance also saturated at fewer resolutions. However, at 2ft resolution, the multiresolution strategies performed as well as or better than single resolution strategies at 1 ft. In our studies we kept the range resolution fixed at 1 ft. This allowed us to explore the collection aperture implications of the multiresolution study.

We used the multiresolution detection strategies as a substitute for the first stage of the Strategic Target Algorithm Research (STAR) algorithm which we have used as a means of detection/discrimination of extended targets in clutter. We applied the

multiresolution approaches to Lincoln Laboratory ADTS data and ERIM DCS data encompassing both targets and clutter. We found that the multiresolution approach provided significantly better detectability of targets (.8 vs. .55) and better false alarm performance (4 vs. 6 false alarms/ km^2). These results were consistent with our TCR findings. We found that using the multiresolution detection schemes, an average 4 dB target to clutter ratio gain was obtained when the DCS and ADTS target/clutter data sets were examined. The multiresolution GLRT provided the best results in our study. We also found that the multiresolution detectors performance was surprisingly robust to starting resolution out to 3ft. Significant detectability gains were encountered there (.75 vs. .48).

We found that the multiresolution detectors also had a much larger pixel-by-pixel detection performance than the single resolution scheme for collected extended targets. This manifests itself in spatial features that can be used for target discrimination. We constructed a tree structured classifier as the basis of our discrimination algorithm. We applied the target detections and false alarms provided by the single resolution and multiresolution detection schemes. We found that the discrimination algorithm provided much better performance based on the multiresolution data than that of the single resolution data. Using the multiresolution schemes, no targets were rejected as false alarms and 273 of 287 false alarms were rejected. For the single resolution case, only 169 of 297 targets were detected with 20 false alarms. This implies that multiresolution processing may provide spatial features that provide high performance discrimination capability.

The appendices have a detailed derivation of a multiresolution sampling strategy for a generalized matched filter detector. Solutions for specific SAR impulse responses are provided. This analysis shows that a significant performance gain can be obtained over single resolution and dyadic multiresolution strategies (wavelets). This analysis would be used as a precursor to a multiresolution target classification algorithm.

Appendix A Resolution Sampling Strategy

In the multiresolution problem, the objective is to use images of various resolutions in order to enhance target detectability. A natural first question is:

- given a known target or a known class of targets, what are the optimal resolutions to be using in order to maximize the performance of your detector?

The purpose of this appendix is to provide an analytical solution to this problem under the assumption of a completely known target signature. In fact we actually solve a more general problem in that we derive a strategy which is optimum when taking into account target detectability and the costs of processing the images, i.e., every new resolution requires some new processing to form an image at that resolution and our general solution takes this into account also.

As stated earlier, there has been a great deal of work on understanding the wavelet transform as a deterministic operator on square-integrable functions and to our knowledge very little work has been done on wavelet transforms in a stochastic environment, which is what we have. Because of our framework, it would be very interesting to research properties of wavelets in this more general stochastic setting.

We make some further assumptions on the clutter process \dot{W} and thermal noise process $N_o(\cdot; \rho)$, in that:

(a1) \dot{W} is a white circular complex Gaussian spatial process in \mathbf{R}^2 with intensity σ_c^2 , i.e., for disjoint bounded sets $A_1, \dots, A_k \subset \mathbf{R}^2$,

$$\left\{ \int_{A_j} \dot{W}(y) dy : 1 \leq j \leq k \right\} \quad (\text{A} - 1)$$

are independent circular complex Gaussian random variables with

$$E \left(\left| \int_{A_j} \dot{W}(y) dy \right|^2 \right) = \sigma_c^2 \text{area}(A_j) \quad 1 \leq j \leq k. \quad (\text{A} - 2)$$

(a2) for fixed ρ , $N_o(\cdot; \rho)$ is a spatially white circular complex Gaussian process.

(a3) for a fixed x the process $N_o(x; \cdot)$ is a circular complex Gaussian process with

$$E(N_o(x; \rho)N_o^*(x; \rho')) = \frac{\sigma_{N_o}^2}{\max\{\rho, \rho'\}} \quad (\text{A} - 3)$$

(a4) \dot{W} and N_o are independent of each other.

For real stochastic processes on an interval T_o of a positive time axis, there is a notion of a Brownian motion process with parameter σ^2 . A real stochastic process $\{X(t) : t \in T_o\}$ is said to be a Brownian motion with scale parameter σ^2 if it is a mean 0 Gaussian process with stationary independent increments and $E(X^2(t)) = \sigma^2 t$ for every $t \in T_o$. Note that in the case of X being a Brownian motion, have

$$E(X(s)X(t)) = \sigma^2 \min\{s, t\} \quad (\text{A} - 4)$$

and in fact this characterizes a real Brownian motion process when it is assumed that the process has mean 0. We now want to generalize the notion of a Brownian motion process to the case where the process is a complex mean 0 Gaussian process. In this case we say a complex stochastic process $\{Y(t) : t \in T_o\}$ with index set $T_o \subset [0, \infty)$ is a complex Brownian motion process with parameter σ^2 if it is a circular complex Gaussian process with the real and imaginary parts of Y being independent Brownian motions with common parameter $\frac{\sigma^2}{2}$.

The assumption in (a3) is essentially equivalent to the assumption that N_o is a complex Brownian motion when the index set is inverted in ρ . More specifically if we redefine the index set of $N_o(\rho)$ to be $\{\frac{1}{\rho} : \rho \in [\rho_l, \rho_u]\}$ and define the process N_r on this index set as

$$N_r\left(\frac{1}{\rho}\right) = N_o(\rho) \quad (\text{A} - 5)$$

then N_r is a complex Brownian motion. This is the essence of the assumption (a3) and this assumption physically is related to the fact that the amount of thermal noise is proportional to the length of the aperture and this is inversely proportional to resolution.

Now based on the above model, we want to derive the optimal sampling (in resolution) strategy for discriminating between the hypothesis H_o and H_1 , or if we think of $I(x; \rho)$ as a stochastic wavelet transform plus noise, then this amounts to choosing the optimal scale parameters of the noisy wavelet transform process. The approach we have taken, because of the complexity of the problems is to fix the location (or known in wavelet jargon as the translation parameter) x which we can assume is 0, and derive the optimal sampling strategy of the noisy wavelet transform process $I(\rho, x)$ in terms of ρ . Of course this is not a fully realistic solution of the original problem (and the astute reader will also have already noted other unrealistic simplifications which will be discussed in more detail in the Summary). But we hope by solving this problem, to be able to shed some light on the analytical choice of the sampling strategy and to provide some light on setting up and solving the much more complex general problem of considering all of the pixel locations simultaneously.

As stated earlier, for the sake of notational simplicity we will assume without loss of generality that the fixed location $x = 0$ and thus we are trying to derive a sampling strategy for discriminating between H_o and H_1 which are no longer a function of x . In doing this we will be using ideas from Cambanis/Masry (1983), hereafter referred to as CM. The new hypotheses based on this fixed location assumption are given by

$$\begin{aligned} H_o \text{ (clutter only)} : I(\rho) &= \frac{1}{\sqrt{\rho}} \int h\left(\frac{y}{\rho}\right) \dot{W}(y) dy + N_o(\rho) \\ &= N(\rho) \end{aligned} \quad (\text{A-6})$$

and

$$\begin{aligned} H_1 \text{ (target present)} : I(\rho) &= \frac{1}{\sqrt{\rho}} \int h\left(\frac{y}{\rho}\right) (\dot{W}(y) + g_t(dy)) dy + N_o(\rho) \\ &= \frac{1}{\sqrt{\rho}} \int h\left(\frac{y}{\rho}\right) g_t(y) dy + \left[\frac{1}{\sqrt{\rho}} \int h\left(\frac{y}{\rho}\right) \dot{W}(y) dy + N_o(\rho) \right] \\ &= S(\rho) + N(\rho) \end{aligned} \quad (\text{A-7})$$

where we have actually substituted $h(-)$ in place of h in the above equations. This was done since it simplifies notation and it makes absolutely no difference to the final

results. Also recall that the index set for the process in both (A-6) and (A-7) are resolutions ρ in the range $[\rho_l, \rho_u]$. By the above representations, we see that N is a circular complex Gaussian process with the autocorrelation function

$$\begin{aligned}
R(\rho, \rho') &= E(N(\rho)N^*(\rho')) \\
&= E \left[\left(\frac{1}{\sqrt{\rho}} \int h \left(\frac{y}{\rho} \right) \dot{W}(y) dy + N_o(\rho) \right) \left(\frac{1}{\sqrt{\rho'}} \int h \left(\frac{y'}{\rho'} \right) \dot{W}(y') dy' + N_o(\rho') \right)^* \right] \\
&= \frac{\sigma_c^2}{\sqrt{\rho\rho'}} \int h \left(\frac{y}{\rho} \right) h^* \left(\frac{y}{\rho'} \right) dy + \frac{\sigma_{N_o}^2}{\max\{\rho, \rho'\}} \quad (A-8)
\end{aligned}$$

This is also the autocovariance function of the process $S(\rho) + N(\rho)$ under the assumption of H_1 . The above formulation is quite similar to the problem considered in CM (1983) with the only differences being that instead of having complex Gaussian processes, they considered only real Gaussian processes and they made a technical assumption on the signal process S relative to the autocorrelation function R . Specifically they made the following assumption:

(a5) the signal process S satisfies: \exists a square integrable function f over $[\rho_l, \rho_u]$ such that

$$\int_{\rho_l}^{\rho_u} R(\rho, \rho') f(\rho') d\rho' = S(\rho). \quad (A-9)$$

The assumption (a5) (and previous assumptions), imply that even if one knew I over the whole interval $[\rho_l, \rho_u]$, it is not possible to perfectly discriminate between H_o and H_1 , i.e., between clutter only and clutter plus target. It also implies that the optimal detector (for a target being present) using values of I across all resolutions in the range of $[\rho_o, \rho_1]$ is given by

$$\text{Accept } H_1 \text{ and declare a target if } \Re \left(\int_{\rho_o}^{\rho_1} I(\rho) f^*(\rho) d\rho \right) > T \quad (A-10)$$

where T is a threshold and \Re denotes the real part. The above is stated in Cambanis/Masry (1983) for the real case and is a straightforward corollary of Karhunen-Loeve expansion (discussed in Appendix A.4) and some analysis using the expansion

to determine the likelihood ratio. This is done in detail in Grenander (1980) (cf. chapter 5). One can think of the test in (A-10) as a matched filter detector.

Of course, in our case we do not have a continuum of data, but rather we must choose a finite number of resolutions. This motivates the determination of a sampling strategy for the resolutions which gives detection performance as close as possible to the performance of the detector in (A-10). Before specifying such a sampling strategy, we mention an interesting property of the stochastic process $\{I(\rho) : \rho_o \leq \rho \leq \rho_1\}$ which is true given some realistic assumptions on the impulse response function h (for example these properties are true if h or its Fourier transform is a sinc function). The property is that under certain conditions on the mother wavelet function h , have:

I is a Markov process in both the positive and negative directions of ρ , i.e., the conditional distribution of $I(\rho_n)$ given $I(\rho_1), \dots, I(\rho_{n-1})$ only depends on $I(\rho_{n-1})$ in the cases of

$$\rho_1 \leq \rho_2 \leq \dots \leq \rho_n \quad (\text{A} - 11)$$

or

$$\rho_n \leq \rho_{n-1} \leq \dots \leq \rho_1. \quad (\text{A} - 12)$$

The latter ordering is probably the more natural ordering to consider since the Markov relationship here is saying that if one has a sequence of images where the resolution is becoming finer and finer, then the conditional distribution of an even finer resolution image is only dependent on the finest resolution previously considered. The technical assumptions on h are outlined rigorously in Appendix A.4 and also given there is a careful derivation and discussion of the Markov property (P). Note that this suggests an alternative framework from which to investigate, namely that of optimal sampling of Markov processes for the purposes of detection. This would be an interesting approach which is significantly different from the approach presented in this appendix.

A.1 First Sampling Strategy

In this subsection, we set up a framework for specifying what we mean by optimal resolutions from a target detection viewpoint. We then use this framework and some analysis to develop a numerical algorithm for determining the optimal resolutions.

To set up the framework consider a fixed number of resolutions, say n , and we suppose the objective is to find the optimal resolutions $\rho_1, \rho_2, \dots, \rho_n$ to choose so that based on $I(\rho_1), \dots, I(\rho_n)$, we would have *optimal* detection performance. Before proceeding, we have to clearly specify what is meant by the term *optimal*. The framework we will use is as follows. We specify the probability of false alarm to be some fixed value which we denote by P_{FA} . For any finite set of resolutions ρ_1, \dots, ρ_n , there is some achievable probability detection given the probability of false alarm is as specified. We denote this probability of detection by $P_D(\mathcal{L}_n)$ where $\mathcal{L}_n = (\rho_1, \dots, \rho_n)$. Mathematically we want to find the resolutions, i.e., the vector \mathcal{L}_n^o which satisfies that

$$P_D(\mathcal{L}_n^o) = \max_{\mathcal{L}_n} P_D(\mathcal{L}_n) \quad (\text{A} - 13)$$

We denote the righthand side (RHS) of (A-13) by P_D^n . As discussed in CM (1983), based on assumption in (a5), there is an upper bound on the probability detection P_D^n which is strictly below 1. It is essentially the probability of detection for optimal test given all the resolutions. As discussed earlier, this test would be based on thresholding the real part of the statistic $\int I(\rho) f^*(\rho) d\rho$ where f is the function given in (a5). We denote the maximal probability of detection for this test by P_D^\dagger . As shown in CM (1983), as n increases, $P_D^n \rightarrow P_D^\dagger$. It turns out that the determination of \mathcal{L}_n^o which satisfies (A-13) is exactly equivalent to another criteria, namely that of finding the resolutions which maximize what is known as the generalized signal-to-noise ratio. This is discussed in greater detail in Appendix C.2, which is essentially results from CM (1983) adapted to our situation. For a set of resolutions $\mathcal{L}_n = (\rho_1, \dots, \rho_n)^t$, we

will denote the GSNR by $G(\mathcal{L}_n)$ and it is defined by

$$G^2(\mathcal{L}_n) = \frac{[E_1(\eta(I(\mathcal{L}_n))) - E_o(\eta(I(\mathcal{L}_n)))]^2}{\text{Var}_o(\eta(I(\mathcal{L}_n)))} \quad (\text{A} - 14)$$

where E_j is the expectation operator under H_j , $j = 0, 1$, Var_o is the variance operator under H_o , and $I(\mathcal{L}_n)$ is the vector $[I(\rho_1), \dots, I(\rho_n)]$, the term $\eta(I(\mathcal{L}_n))$ corresponds to the log of the likelihood ratio (after dropping terms which do not depend on $I(\mathcal{L}_n)$) and is given by

$$\eta(I(\mathcal{L}_n)) = \Re \left((I^*(\mathcal{L}_n))^t \Sigma^{-1}(\mathcal{L}_n) S(\mathcal{L}_n) \right) \quad (\text{A} - 15)$$

where $\Sigma(\mathcal{L}_n)$ is the covariance matrix of the random vector $[N(\rho_1), \dots, N(\rho_n)]^t$ (which is the same under both H_o and H_1), i.e.,

$$(\Sigma(\mathcal{L}_n))_{ij} = E(N(\rho_i)N^*(\rho_j)) \quad 1 \leq i, j \leq n \quad (\text{A} - 16)$$

The interpretation of the GSNR is heuristically simple. It is simply a measure of the separation of the distributions of the log-likelihood ratio $\eta(I(\mathcal{L}_n))$ under H_1 and H_o . Based on the maximization of the GSNR being essentially equivalent to the maximization of the probability of detection, we take an analogous approach to that taken in CM (1983), where they chose \mathcal{L}_n to maximize the GSNR. It is useful to have an alternative interpretation of the resolutions which maximize the GSNR. By substitution of (A-15) into (A-14), it is easy to see that the GSNR has a form given by

$$G^2(\mathcal{L}_n) = (S^*(\mathcal{L}_n))^t \Sigma^{-1}(\mathcal{L}_n) S(\mathcal{L}_n) \quad (\text{A} - 17)$$

where $S(\mathcal{L}_n)$ denotes the vector of $[S(\rho_1), \dots, S(\rho_n)]^t$. The RHS is a quadratic form where the matrix $\Sigma^{-1}(\mathcal{L}_n)$ is positive definite. Hence it turns out that one can think of the RHS of (A-17) as an inner product of $S(\mathcal{L}_n)$ with itself as long as you think of the right inner product. This inner product is actually defined in Appendix A.4. Even more important than this we can think of $S(\mathcal{L}_n)$ as a projection of the continuous

function S in this Hilbert space. Thus we can write the GSNR in (A-17) according to the following expression:

$$G^2(\mathcal{L}_n) = \|\mathcal{P}_{\mathcal{L}_n} S\|_R^2 \quad (\text{A} - 18)$$

where $\|\cdot\|_R$ is the norm in this more general Hilbert space and $\mathcal{P}_{\mathcal{L}_n}$ is a projection operator. For notational convenience we now drop the subscript R on the inner products and norms, since all such refer to this special Hilbert space. This Hilbert space is known as the Reproducing Kernel Hilbert Space (RKHS) and the nicety of it is that evaluations of the process at a time instant corresponds to a projection in this Hilbert space. For the sake of intuition it is entirely appropriate to think of the projections we will be describing as occurring in finite dimensional space. In Appendix A.4, we give a more thorough discussion and background on RKHS's and their applications.

Based on (A-18) and previously quoted results (outlined in Appendix C.2), one optimal algorithm (for resolution selection) proceeds by selecting $\mathcal{L}_n^o = (\rho_{n1}^o, \dots, \rho_{nn}^o)$ as the n resolutions which satisfies

$$\begin{aligned} G^2(\mathcal{L}_n^o) &= \sup_{\mathcal{L}_n} G^2(\mathcal{L}_n) \\ &= \sup_{\mathcal{L}_n} (S^*(\mathcal{L}_n))^t \Sigma^{-1}(\mathcal{L}_n) S(\mathcal{L}_n). \end{aligned} \quad (\text{A-19})$$

The above maximization is over \mathcal{L}_n , i.e., over n -dimensional space and in every iteration of this maximization, one must compute an inverse. Both of these factors suggests that for moderate to large values of n , the above maximization may be difficult to carry out numerically. This provided motivation in CM (1983) to consider alternative schemes for selecting resolutions which, though not optimal for any finite number of resolutions n , are optimal in an asymptotic sense. A number of schemes were presented depending on the properties of the autocorrelation function R , and one of the cases they considered included the case of noise process being an independent increments mean 0 Gaussian process. As proven in Appendix A.4, if we assume

that h is sinc function, our noise process corresponds to a circular complex Gaussian process with independent increments, but not stationary increments. We conjecture that the sampling results presented in CM (1983) have very close analogous results in this case. Thus we present the results in CM (1983) suitably modified for our case as a heuristic algorithm which is likely to have some similar asymptotic optimality properties as stated in CM (1983), though some further research is needed here to verify that this extension is really valid. The procedure is to choose a probability density function ψ whose support is in the interval $[\rho_l, \rho_u]$. Then at stage n , we select n quantile values $\{\rho_{n1}^1, \dots, \rho_{nn}^1\}$ where

$$\int_{\rho_l}^{\rho_{nj}^1} \psi(t) dt = \frac{j-1}{n} \quad 1 \leq j \leq n. \quad (\text{A} - 20)$$

The result proven in CM (1983) (for the real case) is that

$$n^2(P_D^\dagger - P_D^n) \rightarrow K \int_{\rho_l}^{\rho_u} \frac{\beta(\rho)|f(\rho)|^2}{\psi^2(\rho)} d\rho \quad (\text{A} - 21)$$

where K is a known constant and β is a known function related to the autocorrelation function R . In CM (1983), they derived a sufficient condition for the selection procedure to be asymptotically optimal and this condition was that ψ be proportional to $(\beta|f|^2)^{\frac{1}{3}}$.

As stated earlier, we conjecture that a result very close to the above is in fact true, and so we view the above as a useful heuristic for deriving a simpler resolution selection algorithm. One technical snag is that it may be difficult to identify exactly the function f which satisfies the integral equation given by

$$\int R(\rho, \rho') f(\rho') d\rho' = S(\rho). \quad (\text{A} - 22)$$

But since we only want to know the approximate shape of f in order to derive a ψ which is approximately optimal (i.e., select ψ proportional to $(\beta|f|^2)^{\frac{1}{3}}$), it probably

is useful to find a numerical approximation for f given our knowledge of R and S . This we can easily do by discretizing the problem, i.e., choose ρ_1, \dots, ρ_n where

$$\rho_j = \rho_l + \frac{j(\rho_u - \rho_l)}{n} \quad 1 \leq j \leq n \quad (\text{A} - 23)$$

and then noting that (A-22) after discretization can be reformulated as

$$\frac{\rho_l - \rho_u}{n} \begin{bmatrix} R(\rho_1, \rho_1) & \cdots & R(\rho_1, \rho_n) \\ \vdots & \ddots & \vdots \\ R(\rho_n, \rho_1) & \cdots & R(\rho_n, \rho_n) \end{bmatrix} \begin{bmatrix} f(\rho_1) \\ \vdots \\ f(\rho_n) \end{bmatrix} \approx \begin{bmatrix} S(\rho_1) \\ \vdots \\ S(\rho_n) \end{bmatrix}. \quad (\text{A} - 24)$$

Now one just solves this equation numerically for $f(\rho_1), \dots, f(\rho_n)$ and then use these values and some numerical fitting routine for determining an approximation to f .

A.2 General Sampling Strategy

In this section we generalize the sampling strategy to take into account the processing costs. We assume that the costs of processing are linearly inversely related to resolution. However it should be mentioned that our discussion could be carried over to a more general class of loss functions (than linear in the inverse of resolution). To set up the framework we need some notation. For the n -dimensional vector of resolutions, \mathcal{L}_n , we denote the linear span of the autocorrelation functions $\{R(\rho_j, \cdot); 1 \leq j \leq n\}$ by $\mathbf{L}(\mathcal{L}_n)$. We now assume for a fixed n that we want to find resolutions \mathcal{L}_n which minimize

$$\begin{aligned} & K'(\|S\|^2 - \|\mathcal{P}_{\mathbf{L}(\mathcal{L}_n)} S\|^2) + K \sum_{j=1}^n \frac{1}{\rho_j} \\ &= K'(\|S\|^2 - (S^*(\mathcal{L}_n))^t \Sigma^{-1}(\mathcal{L}_n) S(\mathcal{L}_n)) + K \sum_{j=1}^n \frac{1}{\rho_j} \\ &= L_n(\mathcal{L}_n) \end{aligned} \quad (\text{A-25})$$

or equivalently are trying to maximize

$$K'(S^*(\mathcal{L}_n))^t \Sigma^{-1}(\mathcal{L}_n) S(\mathcal{L}_n) - K \sum_{j=1}^n \frac{1}{\rho_j} \quad (\text{A} - 26)$$

Here K, K' are positive constants which reflect the proportional costs of the two differences. Now the first term of $\|S\|^2 - \|\mathcal{P}_{\mathbf{L}(\mathcal{L}_n)} S\|^2$ converges to 0 at a rate which is proportional to the rate at which $P_D^\dagger - P_D^n$ converges to 0, as was stated in the previous subsection. Because it is easier numerically to handle (than $P_D^\dagger - P_D^n$), we base our loss function on $\|S\|^2 - \|\mathcal{P}_{\mathbf{L}(\mathcal{L}_n)} S\|^2$. Since we can divide the loss function by the constant K' and not change its basic structure (optimal resolutions remain the same), we can without loss of generality assume that $K' = 1$ and so we only need to specify the constant K . Now note that as n gets large the loss function goes to ∞ , so there is actually some optimal number of resolutions n_* and corresponding set of n_* optimal resolutions $\rho_1^*, \dots, \rho_{n_*}^*$ which satisfies that

$$L_n(\rho_1^*, \dots, \rho_{n_*}^*) = \inf_n \inf_{\mathcal{L}_n} L_n(\mathcal{L}_n) \quad (\text{A} - 27)$$

This set of resolutions, which we denote as a vector $\mathcal{L}_{n_*}^*$, then represents the optimal set of resolutions across all possible finite sets of resolutions after taking into account the probability of detection and the costs of processing each image. Now K is related to how one views costs of having a lower probability of detection versus paying for more processing of images, but as stated earlier the first term in the loss function, of $\|S\|^2 - \|\mathcal{P}_{\mathbf{L}(\mathcal{L}_n)} S\|^2$ is asymptotically proportional to the difference $P_D^\dagger - P_D(\mathcal{L}_n)$ which is the maximal probability of detection minus the probability of detection based on using the resolutions ρ_1, \dots, ρ_n . This is described in more detail in Appendix C.2 where the precise proportionality constant is given and this may be useful for choosing the constant K . For this loss function and for $n \in \mathbf{N}$, let $\mathcal{L}_n^o = [\rho_{n1}^o, \dots, \rho_{nn}^o]^t$ be a vector of resolutions which minimizes the loss function $L_n(\mathcal{L}_n)$ as a function of n . Again it should be noted that the numerical determination of \mathcal{L}_n^o can be difficult

for moderate to large values of n for identical reasons as mentioned in the previous subsection. Thus in the context of moderate to large values of n , it would be useful to again derive simpler alternative selection algorithms as was done in the previous subsection for the special case when the wavelet transform process had independent increments. This would be a good area for future research.

A.3 Solution to Resolution Sampling Strategies

Introduction

This section discusses the optimal selection of a subset of resolutions from a multiresolution signal to maximize the detection probability. We will call this a resolution sampling strategy. It is not possible to use every resolution, because without an iterative structure in place, the dimensions of the problem quickly become too large. A solution to this will be discussed later. The organization of this section is as follows: First we will pose the general problem of detecting a known signal in additive noise and discuss the solution. Next, an approach based on the work of Cambanis and Masry (1983) for selecting a discrete set of resolutions at which to form the detection statistic is described. The criteria to evaluate performance of the detector using the subset of resolutions are also provided. Both the continuous and discrete solutions are provided. Then, the multiresolution SAR problem is introduced, and the form of the covariance function (which is required by the solution formulation) is derived. In addition, the Markov property is verified and the remaining conditions required for the derivation of the sampling strategy are developed. An example target is described next, and the sampling strategies using both the continuous and discrete solution strategies are obtained and compared. Finally, the performance of the method based on the Cambanis and Masry approach is compared to other methods of selecting sampling strategies, and the saturation behavior as the number of resolution samples increases is investigated.

Problem Statement and Solution Development

As before, we will denote the complex valued SAR image at pixel location $\underline{x} \in \mathcal{R}^2$ which takes into account resolution ρ by $T(\underline{x}, \rho)$. We would like to consider the problem of detecting a multiresolution signal $T_s(\underline{x}, \rho)$ embedded in multiresolution additive noise $T_c(\underline{x}, \rho)$ at a given location $\underline{x} = \underline{x}_0$ over a given resolution interval $\rho_l < \rho < \rho_u$, i.e.,

$$H_1: T(\underline{x}, \rho) = T_s(\underline{x}, \rho) + T_c(\underline{x}, \rho), \rho_l < \rho < \rho_u$$

$$H_0: T(\underline{x}, \rho) = T_c(\underline{x}, \rho), \rho_l < \rho < \rho_u$$

The covariance of the noise process is denoted $R_{\underline{x}_0}(\rho, \rho')$, and

$$R_{\underline{x}_0}(\rho, \rho') = E[T_c(\underline{x}, \rho) T_c^*(\underline{x}, \rho')] \Big|_{\underline{x} = \underline{x}_0}$$

The optimal test statistic is

$$\varphi = \int_{\rho_l}^{\rho_u} f(x_0, \rho) T(x_0, \rho) d\rho$$

where $f(x_0, \rho)$ is the generalized multiresolution matched filter found by solving the integral equation

$$T_s(x_0, \rho) = \int_{\rho_l}^{\rho_u} R_{x_0}(\rho, \rho') f(x_0, \rho') d\rho'.$$

When the solution is to be found in the discrete domain, the process $T(x_0, \rho)$ is sampled at N points (a method to select the points will be described below) $\{\rho_1, \rho_2, \dots, \rho_N\}$ to form the process

$$\underline{T}(x_0, \rho) = \{T(x_0, \rho_1), T(x_0, \rho_2), \dots, T(x_0, \rho_N)\}^t.$$

The test statistic, denoted φ_N to indicate its dependance on the N sample points, then becomes

$$\varphi_N = \underline{f}^H \underline{T}(x_0, \rho)$$

where

$$\underline{f} = R_{x_0}^{-1}(\rho, \rho') \underline{T}_s(x_0, \rho).$$

and $R_{x_0}^{-1}(k, i), i, j = 1 \dots N$ is the inverse matrix of $R_{x_0}(\rho_i, \rho_j), i, j = 1 \dots N$.

To find the optimal set of $\{\rho_i\}$ directly would entail a large dimensional search over all subsets of N resolutions. Cambanis and Masry (1983) show that if $R_{x_0}(\rho, \rho')$ is not differentiable on the diagonal of $[\rho_l, \rho_u] \times [\rho_l, \rho_u]$, then selecting the N quantiles of a function $h^*(x_0, \rho)$, defined by

$$h^*(x_0, \rho) \propto [\beta(x_0, \rho) f^2(x_0, \rho)]^{\frac{1}{3}}$$

will result in an asymptotically optimal solution for the number of resolutions being used. The function $\beta(x_0, \rho)$ is given by

$$\beta(x_0, \rho) = R_{x_0}'(\rho, \rho - 0) - R_{x_0}'(\rho, \rho + 0)$$

where

$$R_{x_0}'(\rho, \rho') = \frac{\partial}{\partial \rho'} R_{x_0}(\rho, \rho').$$

If the noise is stationary,

$$\beta = \left. \frac{dR_{x_0}}{d\rho} \right|_{0^-} - \left. \frac{dR_{x_0}}{d\rho} \right|_{0^+}.$$

Campanis and Masry also show that two measures of performance are monotonically related: $P_d(\varphi)$, the probability of detection for the detector φ at a fixed false alarm rate α , and $S^2(\varphi)$, the generalized signal to noise ratio of the continuous-resolution optimal detector φ . $S^2(\varphi)$ is defined by

$$S^2(\varphi) = \int_{\rho_l}^{\rho_u} T_s(x_0, \rho) f(x_0, \rho) d\rho$$

This relationship between the two measures is given by

$$P_d(\varphi) = \Phi[S(\varphi) - \Phi^{-1}(1 - \alpha)]$$

where Φ is the error function. In discrete form

$$S^2(\varphi_N) = T_s^H(x_0, \rho) R_{x_0}^{-1}(\rho, \rho') T_s(x_0, \rho) \text{ and}$$

$$P_d(\varphi_N) = \Phi[S(\varphi_N) - \Phi^{-1}(1 - \alpha)]$$

Multiresolution SAR Sampling Problem Solution

For the SAR imagery problem, we can derive the functional form of the covariance function

$R_{x_0}(\rho, \rho')$, where $T_c(x_0, \rho)$ is the multiresolution signature of the clutter after it has passed through the SAR imagery system. In addition, we can show that the process is Markov, and is related by a transformation to a Brownian motion process.

Assuming that the system has impulse response $h(y)$, which is a sinc function, then

$$R_{x_0}(\rho, \rho') = \frac{1}{\sqrt{\rho\rho'}} \int h\left(\frac{y}{\rho}\right) h^*\left(\frac{y}{\rho'}\right) dy$$

We know that the Fourier transform of $h(y)$ is an indicator function, denoted $\hat{h}(u)$, and also that

$$\frac{1}{\rho} h\left(\frac{y}{\rho}\right) \leftrightarrow \hat{h}(\rho u)$$

where the double-sided arrow denotes Fourier transform. So,

$$R_{x_0}(\rho, \rho') = \frac{1}{\sqrt{\rho\rho'}} \int h\left(\frac{y}{\rho}\right) h^*\left(\frac{y}{\rho'}\right) dy = \sqrt{\rho\rho'} \int \frac{1}{\rho} h\left(\frac{y}{\rho}\right) \frac{1}{\rho'} h^*\left(\frac{y}{\rho'}\right) dy$$

and thus by Parseval's Theorem,

$$R_{x_0}(\rho, \rho') = \sqrt{\rho\rho'} \int \hat{h}(\rho u) \hat{h}(\rho' u) du$$

Now, for $\rho < \rho'$ via properties of the indicator function,

$$\begin{aligned} R_{x_0}(\rho, \rho') &= \sqrt{\rho\rho'} \int \hat{h}(\rho' u) du \\ R_{x_0}(\rho, \rho') &= \sqrt{\frac{\rho}{\rho'}} \int \rho' (\hat{h}(\rho' u)) du \\ R_{x_0}(\rho, \rho') &= \sqrt{\frac{\rho}{\rho'}} \int \hat{h}(u') du' \end{aligned}$$

For $\rho > \rho'$

$$\begin{aligned} R_{x_0}(\rho, \rho') &= \sqrt{\rho\rho'} \int \hat{h}(\rho u) du \\ R_{x_0}(\rho, \rho') &= \sqrt{\frac{\rho'}{\rho}} \int \rho (\hat{h}(\rho u)) du \\ R_{x_0}(\rho, \rho') &= \sqrt{\frac{\rho'}{\rho}} \int \hat{h}(u') du' \end{aligned}$$

Now, assume that $\hat{h}(u)$ has unit energy, i.e.,

$$\int |\hat{h}(u)|^2 du = 1$$

and therefore let

$$A = \int \hat{h}(u) du$$

Then, the covariance $R_{x_0}(\rho, \rho')$ is given by

$$R_{x_0}(\rho, \rho') = \begin{cases} A\sqrt{\frac{\rho}{\rho'}}, & \rho < \rho' \\ A, & \rho = \rho' \\ A\sqrt{\frac{\rho'}{\rho}}, & \rho > \rho' \end{cases}$$

Note that $R_{x_0}(\rho, \rho')$ is continuous at $\rho = \rho'$, but it is not differentiable.

To demonstrate that the process is Markov, the correlation function must satisfy a scaling law condition with respect to resolution [Wong and Hajek]. Let $\rho_1 < \rho_2 < \rho_3$. Then,

$$R_{x_0}(\rho_1, \rho_3) = A\sqrt{\frac{\rho_1}{\rho_3}} = \frac{A\sqrt{\frac{\rho_1}{\rho_2}}A\sqrt{\frac{\rho_2}{\rho_3}}}{A\sqrt{\frac{\rho_2}{\rho_2}}} = \frac{R_{x_0}(\rho_1, \rho_2)R_{x_0}(\rho_2, \rho_3)}{R_{x_0}(\rho_2, \rho_2)}$$

Thus, the process is Markov.

To find $\beta(x_0, \rho)$, note that

$$R_{x_0}'(\rho, \rho') = \begin{cases} \frac{(-A)\sqrt{\rho}}{2\sqrt{\rho'}^3}, & \rho < \rho' \\ 0, & \rho = \rho' \\ \frac{A}{2\sqrt{\rho\rho'}}, & \rho > \rho' \end{cases}$$

therefore

$$\beta(x_0, \rho) = -\frac{A}{\rho}$$

Now, we will determine the transformation which maps the process $T_c(x_0, \rho)$ into Brownian motion. Since the process $T_c(x_0, \rho)$ is Markov, its autocorrelation function $R_{x_0}(\rho, \rho')$ satisfies the equation

$$R_{x_0}(u, s) = \frac{R_{x_0}(u, t) R_{x_0}(t, s)}{R_{x_0}(t, t)}, \rho_l \leq s < t < u \leq \rho_u$$

Letting $u = \rho_u$ and rearranging,

$$R_{x_0}(t, s) = \frac{R_{x_0}(\rho_u, s) R_{x_0}(t, t)}{R_{x_0}(\rho_u, t)} = g(s) k(t)$$

where

$$g(s) = R_{x_0}(\rho_u, s) \text{ and}$$

$$k(t) = \frac{R_{x_0}(t, t)}{R_{x_0}(\rho_u, t)}$$

Let τ be defined as

$$\tau(t) = \frac{g(t)}{k^*(t)}$$

and Z as

$$Z(\rho) = k(\rho) Y(\tau(\rho)), \rho \in [\rho_l, \rho_u].$$

where Y is a circular complex Brownian motion process. Then,

$$R_z(\rho, \rho') = R_{x_0}(\rho, \rho')$$

We have described an invertible transformation which generates a process with the same distribution as the one we have been considering from a Brownian motion process. Thus, we can use the inverse transformation on the original process to generate a system equation which is driven by Brownian noise. The integral equation for this system will be easier to solve, and the solution can be transformed back to give a solution to the original problem. The transformation proceeds from

the original equation

$$T(x_0, \rho) = T_s(x_0, \rho) + T_c(x_0, \rho)$$

to

$$\frac{T(x_0, \tau^{-1}(\rho))}{k(\rho)} = \frac{T_s(x_0, \tau^{-1}(\rho))}{k(\rho)} + Y(x_0, \rho)$$

or

$$\hat{T}(x_0, \hat{\rho}) = \hat{T}_s(x_0, \hat{\rho}) + Y(x_0, \hat{\rho})$$

This problem can be solved for $\hat{f}(x_0, \hat{\rho})$ via the integral equation, the optimal sampling strategy $\{\hat{\rho}_i\}$ can be found via the procedure outlined above. This transformed set of resolutions can then be converted to the set $\{\rho_i\}$.

The transformations $k(\rho)$ and $\tau(\rho)$ can be determined using the equations given above, which are defined on the range $\rho_l < \rho, \rho' < \rho_u$ (and the transformed range $\hat{\rho}_l < \rho, \rho' < \hat{\rho}_u$).

$$g(\rho') = R_{x_0}(\hat{\rho}_u, \rho') = A \sqrt{\frac{\rho'}{\hat{\rho}_u}}$$

$$k(\rho) = \frac{R_{x_0}(\rho, \rho)}{R_{x_0}(\hat{\rho}_u, \rho)} = \sqrt{\frac{\hat{\rho}_u}{\rho}}$$

$$\tau(\rho) = \frac{g(\rho)}{k^*(\rho)} = A \frac{\rho}{\hat{\rho}_u}$$

So,

$$\hat{T}_s(x_0, \hat{\rho}) = \frac{s(\tau^{-1}(\rho))}{k(\rho)} = \sqrt{\frac{\rho}{\hat{\rho}_u}} T_s(x_0, \frac{\hat{\rho}_u}{A\rho})$$

Three Point Target Example

It is instructive to attempt to find the generalized matched filter, $f(x_0, \rho)$, and the optimal sampling strategy $\{\rho_i\}$, $i = 1 \dots N$ for a simple target and a given N , and compare the results from the continuous solution to those obtained from the discrete solution. Consider the case of three equal amplitude colinear scatterers at locations x_i , $i = 1 \dots 3$, all at range $y = 0$ and equidistant from each other. The target $t(x, y)$ can be written as

$$t(x, y) = \sum_{i=1}^3 \delta(x - x_i) \delta(y)$$

The multiresolution signature of this target at any position x can be evaluated using the SAR system equation, and is

$$T_s(x, \rho) = \sum_{i=1}^3 e^{i2\pi x_i^2 / \lambda r_a} \text{sinc}\left(\frac{x - x_i}{\rho}\right)$$

The magnitude of the multiresolution target signature is shown in Figure A-1 at $x_0 = \{0, 0\}$. Figure A-2 provides an example of several sample multiresolution clutter signatures for comparison. Both plots show magnitude versus resolution.

In this simple case, the integral equation for the solution $f(x_0, \rho)$ is difficult, if not impossible to solve. We will use the transformation described previously to solve this problem.

After applying the transformation, target representation at $x = 0$ thus becomes

$$\hat{T}_s(x_0, \rho) = \sqrt{\frac{\rho}{\hat{\rho}_u}} \sum_{i=1}^3 e^{i2\pi x_i^2 / \lambda r_a} \text{sinc}\left(\frac{-A\rho x_i}{\hat{\rho}_u}\right)$$

The integral equation becomes

$$\hat{T}_s(x_0, \rho) = \int_{\hat{\rho}_l}^{\hat{\rho}_u} R_Y(\rho, \rho') \hat{f}(x_0, \rho) d\rho'$$

where

$$R_Y(\rho, \rho') = \min(\rho, \rho')$$

We can rewrite the integral equation using the expression for $R_Y(\rho, \rho')$

$$\hat{T}_s(x_0, \rho) = \int_{\hat{\rho}_l}^{\rho} \rho' \hat{f}(x_0, \rho') d\rho' + \rho \int_{\rho}^{\hat{\rho}_u} \hat{f}(x_0, \rho') d\rho'$$

Taking the partial derivative with respect to ρ of both sides gives

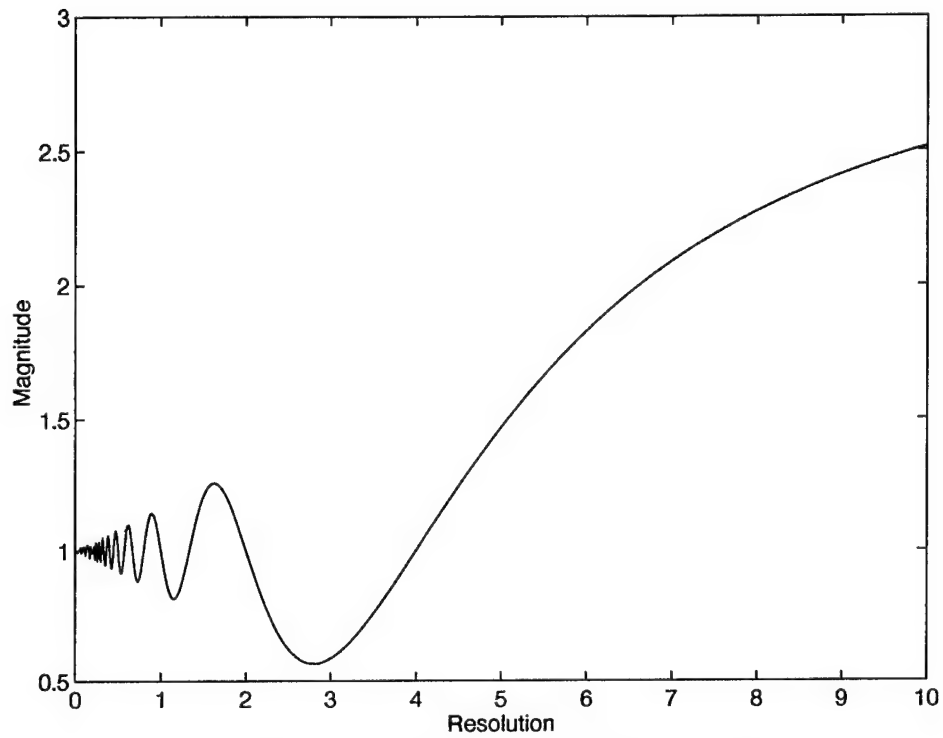


Figure A-1: Target magnitude across resolution.

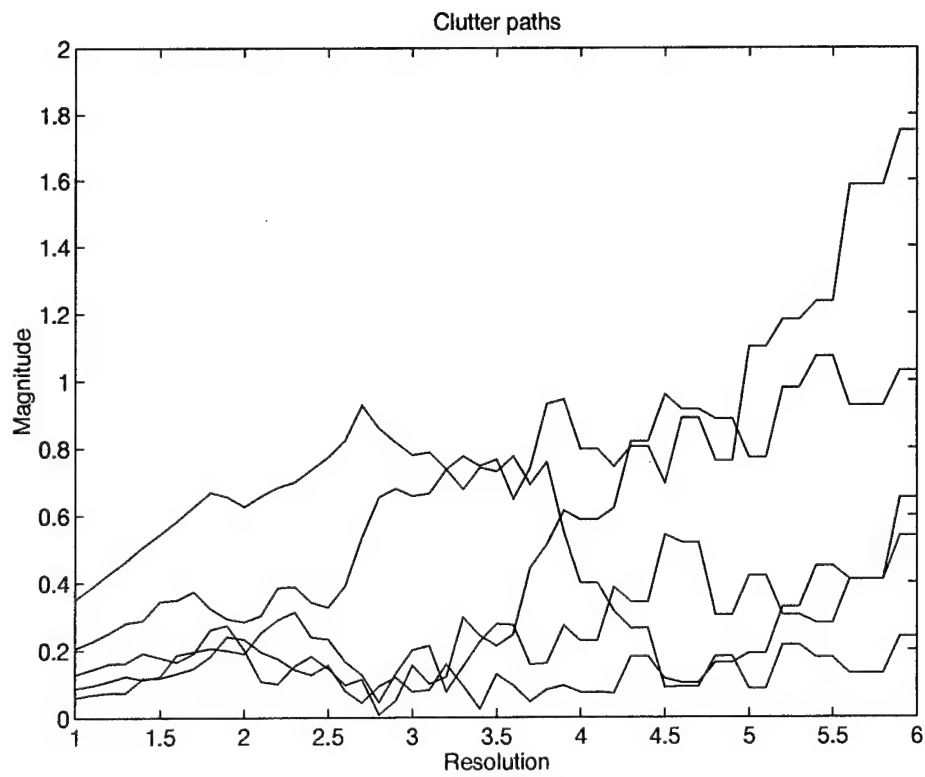


Figure A-2: Examples of clutter magnitude across resolution

$$\frac{\partial}{\partial \rho} \hat{T}_s(x_0, \rho) = \rho \hat{f}(x_0, \rho) + \int_{\rho}^{\hat{\rho}_u} \hat{f}(x_0, \rho') d\rho' - \rho \hat{f}(x_0, \rho) = \int_{\rho}^{\hat{\rho}_u} \hat{f}(x_0, \rho') d\rho'$$

Taking the partial derivative again with respect to ρ yields

$$\frac{\partial^2}{\partial \rho^2} \hat{T}_s(x_0, \rho) = -\hat{f}(x_0, \rho)$$

To complete the solution, the quantiles of

$$\hat{h}^*(x_0, \rho) = \left\| \hat{\beta}(x_0, \rho) \hat{f}^2(x_0, \rho) \right\|^{\frac{1}{3}}$$

where $\hat{\beta}(x_0, \rho) = -1$ (determined from $R_Y(\rho, \rho')$) must be determined. Substituting the results from the solution to the integral equation

$$\hat{h}^*(x_0, \rho) = \left\| \left(\frac{\partial^2}{\partial \rho^2} \hat{T}_s(x_0, \rho) \right)^2 \right\|^{\frac{1}{3}},$$

which can be shown to be

$$\left\| \sum_{i=1}^3 e^{\frac{i2\pi x_i^2}{\lambda r_a}} \left[\frac{\sin\left(\frac{-A\rho x_i}{\hat{\rho}_u}\right)}{\left(\frac{-A\rho x_i}{\hat{\rho}_u}\right)} \left(\frac{-A^2 x_i^2 \sqrt{\rho}}{\hat{\rho}_u^{\frac{5}{2}}} + \frac{1}{2\rho \sqrt{\rho \hat{\rho}_u}} \right) + \frac{\cos\left(\frac{-A\rho x_i}{\hat{\rho}_u}\right)}{\left(\frac{-A\rho x_i}{\hat{\rho}_u}\right)} \frac{A x_i}{\hat{\rho}_u \sqrt{\rho \hat{\rho}_u}} \right] \right\|^2 \right\|^{\frac{1}{3}}$$

The quantiles $\{\hat{\rho}_i\}$ of this function were calculated numerically, then transformed back, and are (1.2, 1.6, 2.3, 3.2, 4.9, 10.0).

For the three point scatterer case, the discrete solution was calculated directly from $T_s(x_0, \rho)$ and $R_{x_0}(\rho, \rho')$ as (1.1, 1.2, 1.6, 2.3, 3.4, 10.0). The GSNR and ROCs were calculated as described above and are shown below in Figure A-3.

Figure A-3 shows the ROC curves (detection versus false alarm probability) for the two solutions described above - exact solution to the three point scatter problem (solution of the integral equation) versus discrete solution (inversion of sampled covariance matrix to determine f). The original sampling density of the discrete covariance matrix included a range of resolutions from 1 to 10, sampled every 0.1. The exact solution does outperform the approximate solution, as was expected.

Comparison of Cambanis/Masry Approach to Other Sampling Strategies

Figure A-4 show the performance of the discrete Cambanis/Masry method to determine a sampling strategy for the three point target problem versus the multiresolution sampling strategy used in the previous project (called multires) and a dyadic splits, or wavelet strategy. The dyadic splits are determined as 1 ft., 2 ft., 4 ft., etc. The number of resolutions ($N = 6$) used for each strategy, and the signal to noise ratio (determined by A) were held constant. Use of the Cambanis/Masry approach to determine the sampling strategy resulted in increased performance, especially over the wavelet sampling approach. Saturation Behavior of Cambanis/Masry Sampling Strategy Approach

Figure A-5 illustrates the saturation behavior of the discrete version of the Cambanis/Masry sampling strategy. A new "optimal" sampling strategy was generated for each value of N , the number of resolution samples used. For each case, the finest resolution used was 1 ft., the coarsest resolution used was 10 ft. For the $N = \text{infinity}$ case, all 91 available samples were used. Little improvement is seen for N 's larger than 12, however there is significant improvement beyond the single resolution case ($N=1$),

Saturation Behavior of Dyadic Split Sampling Strategy Approach

Figure A-6 shows the saturation behavior of the dyadic split/wavelet approach to sampling (the wavelet sampling strategy was fed to the same GSNR/ROC machinery as if was the sampling strategy selected by the C/M mechanism). Performance saturates at lower values of N ($N=6$ versus $N=12$) for the wavelet sampling strategy, and the same levels of performance are not achieved. Note that for this strategy, each increase in N uses a coarser resolution, and little information is gained to aid performance as resolution becomes coarser.

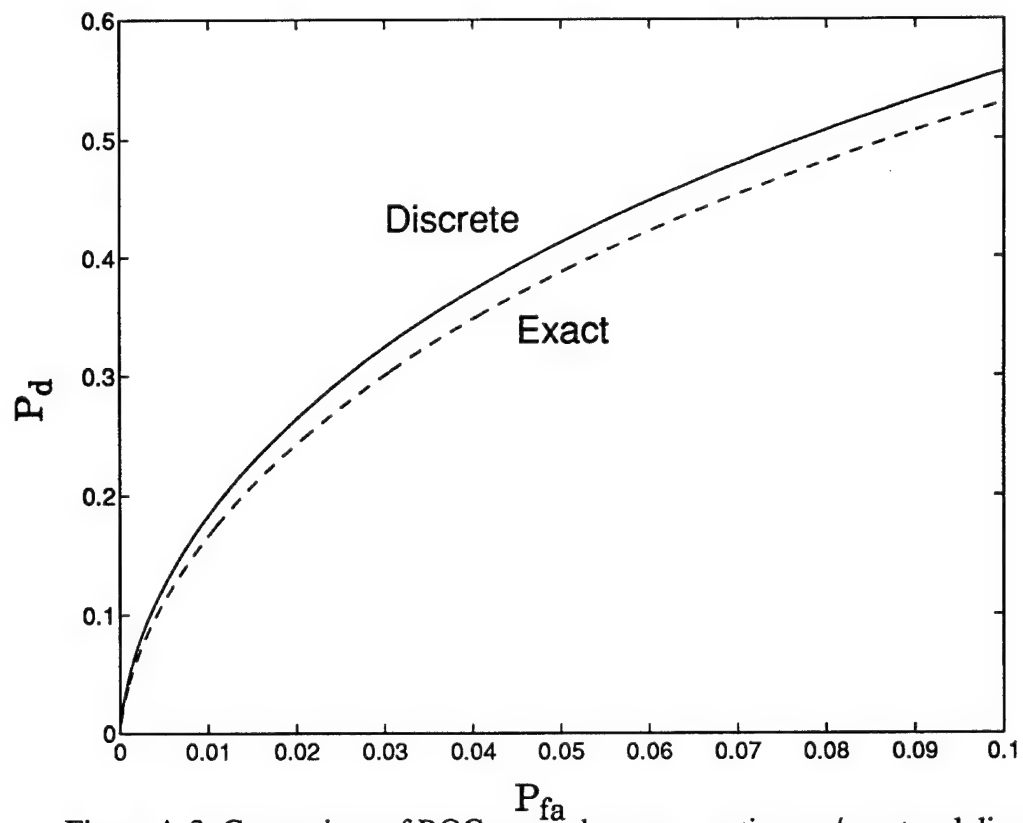


Figure A-3: Comparison of ROC curves between continuous/exact and discrete solutions for the optimal sampling strategy.

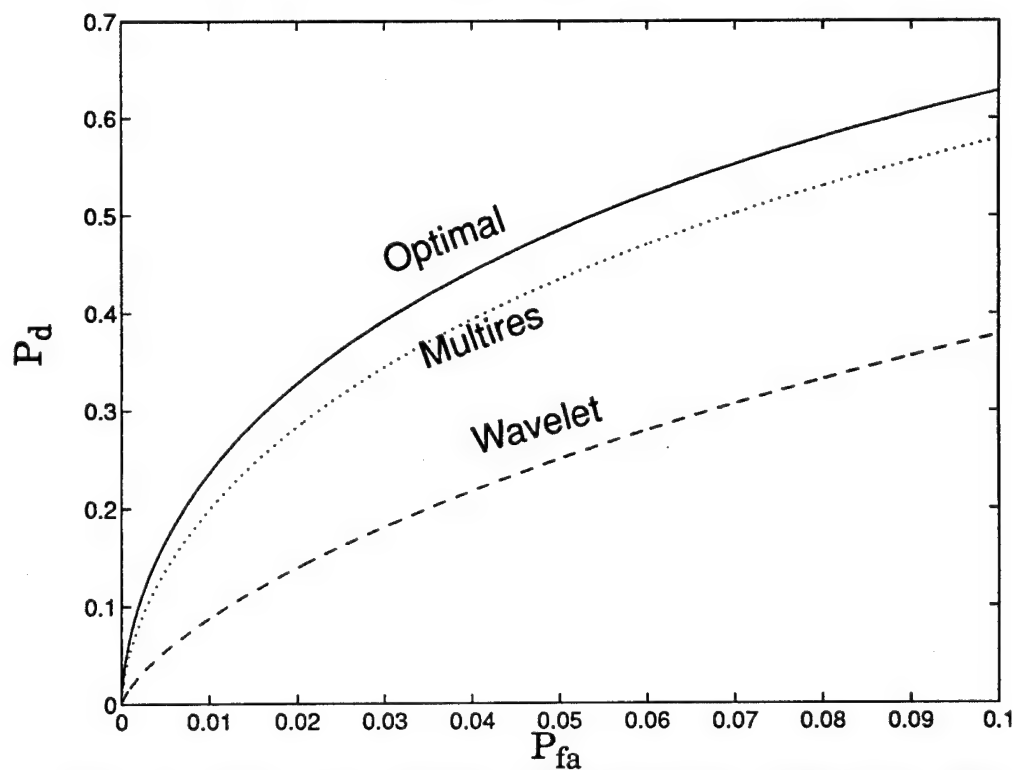


Figure A-4: ROC curves comparing different sampling strategies at a fixed number of samples..

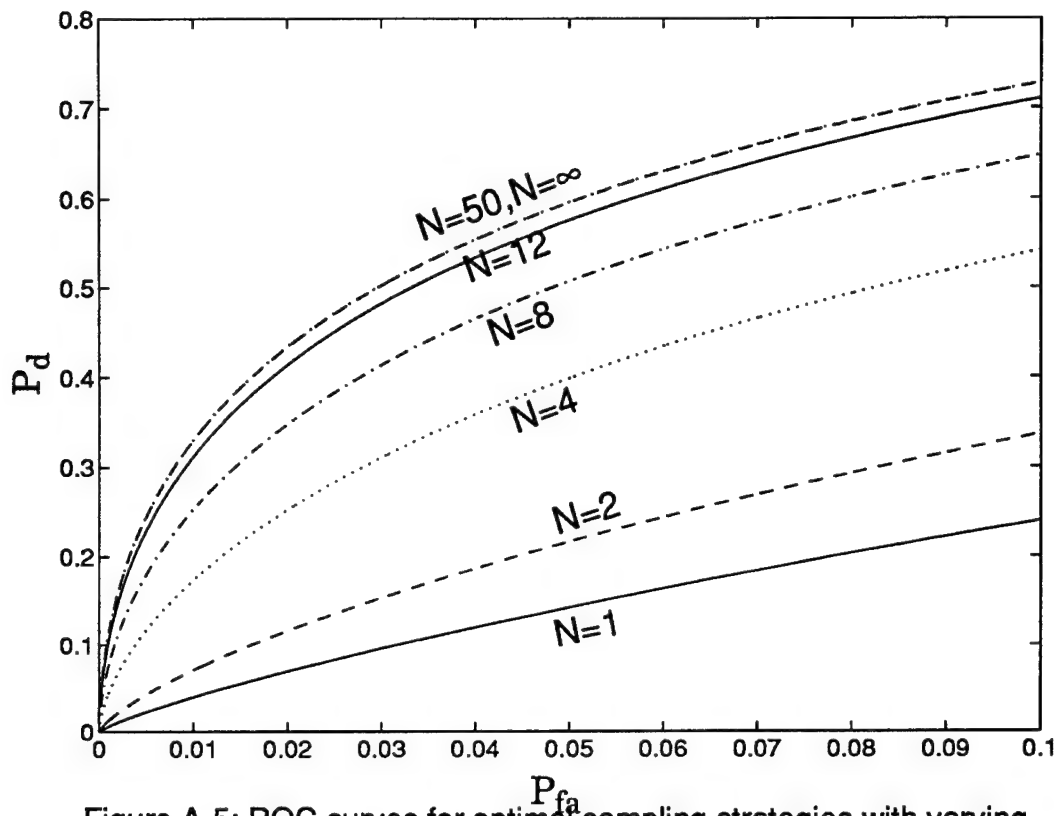


Figure A-5: ROC curves for optimal sampling strategies with varying numbers of samples.

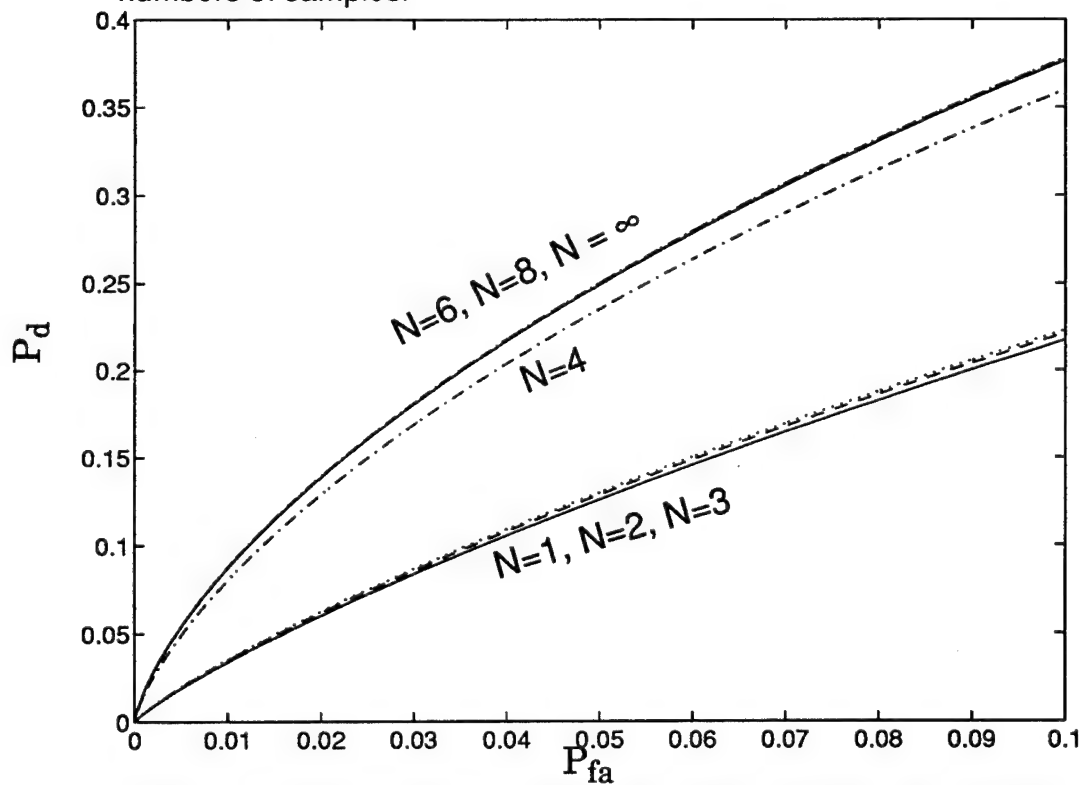


Figure A-6: ROC curves for wavelet sampling strategies with varying numbers of samples.

A.4 Summary

In this appendix we have given a methodology for choosing the optimal resolutions from a target detection criteria. We did this under two cases; the first case specified that the only criteria was to maximize the probability of detection and the second case specified a criteria which included the maximization of the probability of detection *and* a minimization of processing. The technical approach was to model the resolutions from a wavelet transform viewpoint and allowing the scale to be over a continuum and restricting the sampling location to be at a fixed point. We provided an algorithm for determining the optimal resolutions under both criteria and we stated some rigorous properties of these optimal resolutions. We also derived some interesting properties of the wavelet transform process representing our different resolution images.

Recommended areas for future research are:

(i) Generalize the original model to include a random phase on the target, i.e., assume a model where

$$I(x; \rho) = \frac{1}{\sqrt{\rho}} \int h\left(\frac{x-y}{\rho}\right) (\dot{W}(y) + g_t(y)e^{iV}) dy + N_o(x; \rho) \quad x \in \mathcal{X}_\rho \quad (\text{A} - 28)$$

where V is random variable with uniform distribution over $(0, 2\pi)$ and is independent of the clutter process \dot{W} and the thermal noise process N_o . This model is much more realistic since we never expect to know the phase of the target reflectivity (would require precise knowledge of the radar system and collection geometry).

(ii) In the approach in this appendix, we only looked at one pixel location and changing resolutions at that location. The more realistic scenario is to look at optimal resolutions based on considering the data at all pixel locations, i.e., based on

$$\{I(x; \rho) : x \in \mathcal{X}_\rho\} \quad \rho_l \leq \rho \leq \rho_u. \quad (\text{A} - 29)$$

This would mean considering the full wavelet transform both in the scale parameter and in the translation parameter. One of the problems here is that the number of

pixels is different for different resolutions and so there is a problem comparing images of different sizes. A possible starting point for solving this problem is that of breaking up the scene into small square subpatches which are much smaller than the finest resolution, and then approximate $I(x; \rho)$ as a linear combination of these subpatches. This would be a sort of a discrete wavelet transform model-based approach to the multiresolution problem. Another possible approach to this problem would be to look at different metrics relating the probability distributions of two different images at two different resolutions (metrics should be closely related to the probability of detection), which don't need the same number of pixels.

(iii) Investigate simple resolution selection algorithms for the general case of a loss function incorporating the costs of processing. As discussed earlier, there are such algorithms in the case of no processing costs when the general noise process N has independent increments (true if impulse response is a sinc function).

(iii') Derive simple resolution selection algorithms under the framework given in (ii).

(iv) Develop an analogous framework (including simple algorithms) to that presented in the appendix in the case where $P_D^\dagger = 1$. Recall that we used the generalized signal to noise ratio (GSNR) and the justification for doing so was that the resolutions which maximize this (GSNR) also are optimal from a probability detection criteria. But this depended on the assumption of (a5) which implied that $P_D^\dagger < 1$. In many cases (e.g., such as delta functions as part of the target), this does not hold and it would be of interest to research what happens in this more general framework.

Appendix B Properties of Noisy Wavelet Transform

In this appendix, we establish a couple of interesting properties for the I process which were given in Section 2. As stated there, we can view I as a noisy stochastic wavelet transform. The first property we present is that the autocorrelation function R is continuous provided the mother wavelet function h is square-integrable, which we assumed it was. The second result we prove is that under some technical conditions on the function h , the wavelet transform process at a fixed translation is a Gaussian Markov process in the scale parameter. We now state explicitly the first result.

Proposition Appendix B.1 *Suppose h is square-integrable and R is the autocorrelation of the wavelet transform process I . Then R is continuous.*

Proof (sketch). The autocorrelation function is given by

$$R(\rho, \rho') = \frac{\sigma_c^2}{\sqrt{\rho\rho'}} \int h\left(\frac{y}{\rho}\right) h^*\left(\frac{y}{\rho'}\right) dy + \frac{\sigma_{N_o}^2}{\max\{\rho, \rho'\}} \quad (\text{B} - 1)$$

where the second term represents the autocorrelation of the additive thermal noise process and the first term represents the autocorrelation of the integrated clutter process (convolved with scaled version of h). Now to prove the result it suffices to show that the first term is continuous in the argument (ρ, ρ') . This can be done by a very common analysis trick whereby we approximate h by a function h_o which is continuous and has compact support. We then use this trick and the Cauchy-Schwarz inequality to prove the desired result. The details are messy and hence are omitted.

We now state a second result on sufficient conditions on h for ensuring the Markov property for I in both directions and we do this precisely in the form of a Proposition and then we give a proof. But before doing so, we need a preliminary result which

is essentially given in Wong/Hajek (1985), but which we state as a Lemma both for completeness and because we do need to adapt their result to the case of circular complex Gaussian processes.

Lemma Appendix B.2 *Suppose $0 < t_l < t_u < \infty$ and suppose $X = \{X(t) : t_l \leq t \leq t_u\}$ is a circular complex Gaussian process which satisfies that,*

$$R(t, t) > 0 \quad t_l \leq t \leq t_u \quad (\text{B} - 2)$$

and that

$$\mu(\{s : t_l \leq s \leq t_u, R(t, s) = 0\}) = 0 \quad \forall t \quad (\text{B} - 3)$$

where μ is Lebesgue measure (for intuition, notice that this technical condition is satisfied if for every t the function $R(t, s) = 0$ for at most a countable set of s 's).

Then the following conditions are all equivalent:

- (i) X is forward Markov.
- (ii) X is reverse Markov.
- (iii) the autocorrelation function R satisfies that

$$R(u, s) = \frac{R(u, t)R(t, s)}{R(t, t)} \quad t_l \leq s < t < u \leq t_u. \quad (\text{B} - 4)$$

(iv) X has the same distribution as Z where $Z(t) = f(t)Y(\tau(t))$ where f is a deterministic function, Y is circular complex Brownian motion, and τ is non-decreasing function from $[t_l, t_u]$ to $[0, \infty)$.

Proof. By the general theory of Markov processes (cf. Wong/Hajek (1985), pp. 65), it is well-known that (i) and (ii) are equivalent. Thus we only need to verify that (i), (iii), and (iv) are equivalent. Suppose (i) is true and consider $s \leq t \leq u$. Then

$$E(X(u)|X(s)) = \frac{R(u, s)}{R(s, s)}X(s) \quad (\text{B} - 5)$$

by Lemma Appendix B.3 which follows. Also we have that

$$\begin{aligned}
E(X(u)|X(s)) &= E(E(X(u)|X(s), X(t))|X(s)) && \text{prop. of cond. expectations} \\
&= E(E(X(u)|X(t))|X(s)) && \text{by Markov property of } X \\
&= \frac{R(u, t)}{R(t, t)} E(X(t)|X(s)) \\
&= \frac{R(u, t)}{R(t, t)} \frac{R(t, s)}{R(s, s)} X(s). && \text{(B-6)}
\end{aligned}$$

Since $X(s)$ is non-degenerate, (B-5) and (B-6) imply that (iii) holds.

Now suppose (iii) is true. Then we first claim that

$$R(t, s) \neq 0 \quad \text{for every } s, t \quad \text{(B-7)}$$

To see that this is true, suppose the contrary, i.e., suppose that $\exists s < u$ such that

$$R(u, s) = 0. \quad \text{(B-8)}$$

But the hypothesis in (iii) implies that

$$\frac{R(u, t)R(t, s)}{R(t, t)} = 0 \quad \forall t \in (s, u). \quad \text{(B-9)}$$

But this would imply that either $R(u, t) = 0$ or $R(t, s) = 0$ for every $t \in (s, u)$, and it is easy to show that this violates the assumption contained in (B-3) and hence have a contradiction. Thus we have verified the claim expressed by (B-7). Now

$$\begin{aligned}
R(t, s) &= \frac{R(t, u)R(u, s)}{R(u, u)} && t \geq s \\
&= g(s)f(t) && \text{(B-10)}
\end{aligned}$$

where

$$g(s) = R(t_u, s) \quad \text{(B-11)}$$

and

$$f(t) = \frac{R(t, t)}{R(t_u, t)} \quad (\text{B} - 12)$$

Now let τ be defined by

$$\tau(t) = \frac{g(t)}{f^*(t)} \quad t \in [t_l, t_u] \quad (\text{B} - 13)$$

Clearly τ is real and positive since

$$g(t)f(t) = R(t, t) > 0 \quad \forall t. \quad (\text{B} - 14)$$

Finally τ is non-decreasing since for $s \leq t$, have

$$\begin{aligned} \tau(s) &= \frac{g(s)}{f^*(s)} \\ &= \frac{g(s)f(t)}{f^*(s)f(t)} \\ &= \frac{R(t, s)}{f(s)f(t)} \\ &\leq \sqrt{\frac{R(t, t)R(s, s)}{(f^*(s)f(t))^2}} \quad \text{by Cauchy-Schwarz ineq.} \\ &= \sqrt{\frac{g(t)f(t)g^*(s)f^*(s)}{(f^*(s)f(t))^2}} \\ &= \sqrt{\frac{g(s)}{f(s)}} \sqrt{\frac{g(t)}{f(t)}} \\ &= \sqrt{\tau(s)\tau(t)}. \end{aligned} \quad (\text{B-15})$$

Now let

$$Z(t) = f(t)Y(\tau(t)) \quad t \in [t_l, t_u] \quad (\text{B} - 16)$$

Clearly Z is a circular complex Gaussian process and for $s \leq t$,

$$E(Z^*(s)Z(t)) = f^*(s)f(t)E(Y^*(\tau(s))Y(\tau(t)))$$

$$\begin{aligned}
&= f^*(s)f(t)\tau(s) \\
&= g(s)f(t) \\
&= R(t,s)
\end{aligned} \tag{B-17}$$

Hence (iv) has been verified.

That (iv) implies (i) is immediate since Y has independent increments and so automatically is Markov, and by recalling that f is a function which never vanishes (i.e., never equals 0).

Lemma Appendix B.3 *Suppose (U, V) is \mathbb{C}^2 -valued random vector which is circular complex Gaussian with $R_{VU} = E(VU^*)$ and $R_{UU} = E(UU^*)$. Then*

(a) *the random vector*

$$V - \frac{R_{VU}}{R_{UU}}U \tag{B-18}$$

is independent of U .

(b) *the conditional expectation of V given U satisfies that*

$$E(V|U) = \frac{R_{VU}}{R_{UU}}U. \tag{B-19}$$

Proof. (b) is an easy consequence of (a) and (a) is well-known result obtained by recognizing that any linear combination of (U, V) is again circular complex Gaussian.

We now state a couple of lemmas and give a definition before giving the main result (Proposition Appendix B.6) of this appendix, which are sufficient conditions for the wavelet transform process to be Markovian.

Lemma Appendix B.4 *Suppose X and Y are two independent Markov processes. Then $X + Y$ is Markov.*

Proof. Easy consequence of the properties of conditional expectation.

Lemma Appendix B.5 Suppose $\{X(t) : 0 \leq t_l \leq t \leq t_u < \infty\}$ is a circular complex Gaussian process which is Markov and whose autocorrelation function R satisfies that

$$R(t, s) = R(s, s) > 0 \quad s \leq t. \quad (\text{B} - 20)$$

Then there exists non-decreasing function η from $[t_l, t_u]$ to $[0, \infty)$ such that X has same distribution as the process $\{Y(\tau(t)) : t_l \leq t \leq t_u\}$ where Y is a complex Brownian motion process and in particular X has independent increments.

Proof. This is a straightforward corollary to the proof of Lemma Appendix B.2. Specifically it suffices to show that the function f in that proof is 1. But recall that

$$\begin{aligned} f(t) &= \frac{R(t, t)}{R(t_u, t)} && \text{by def. of } f \text{ in proof} \\ &= \frac{R(t, t)}{R(t, t)} && \text{by the hypothesis} \end{aligned}$$

Definition/Notation. For square integrable function f from \mathbb{R}^2 to \mathbb{C} , we denote the Fourier transform by \hat{f} . We let \mathcal{H} denote a special class functions defined by $h \in \mathcal{H}$ if and only

$$h(y_1, y_2) = C|y_1|^{k_1}|y_2|^{k_2}1_{[\alpha_1, \beta_1]}(y_1)1_{[\alpha_2, \beta_2]}(y_2) \quad y_1, y_2 \in \mathbb{R}. \quad (\text{B} - 21)$$

where $C > 0$, $k_1, k_2 \geq 0$ and $\alpha_1 < 0 < \beta_1$ and $\alpha_2 < 0 < \beta_2$, and

$$1_{[\alpha_j, \beta_j]}(y_j) = \begin{cases} 1 & y_j \in [\alpha_j, \beta_j] \\ 0 & \text{otherwise} \end{cases} \quad j = 1, 2. \quad (\text{B} - 22)$$

Proposition Appendix B.6 Suppose have the wavelet transform process I (under H_o) where

$$I(\rho) = \int \frac{1}{\sqrt{\rho}} h\left(\frac{y}{\rho}\right) (\dot{W}(y) + g_t(y)) dy + N_o(\rho) \quad (\text{B} - 23)$$

where \dot{W} is Gaussian white noise with parameter σ_c^2 and N_o is a circular complex Gaussian process. Assume the following:

- (i) \dot{W} and N_o are independent.
- (ii) the process $N_o(x : \cdot)$ is a circular complex Gaussian process with

$$E(N_o(x; \rho) N_o^*(x; \rho')) = \frac{\sigma_{N_o}^2}{\max\{\rho, \rho'\}} \quad \rho_l \leq \rho, \rho' \leq \rho_u. \quad (\text{B} - 24)$$

- (iii) either $h \in \mathcal{H}$ or $\hat{h} \in \mathcal{H}$.

Then :

- (a) I is a Markov process in both the positive and negative directions.
- (b) in the special case of either h or \hat{h} are equal to $C1_{[\alpha_1, \beta_1]}(y_1)1_{[\alpha_2, \beta_2]}(y_2)$, then I has independent increments.

Proof. Let V be the process given by

$$V(\rho) = \int h\left(\frac{y}{\rho}\right) \dot{W}(y) dy \quad \rho_l \leq \rho \leq \rho_u. \quad (\text{B} - 25)$$

By the independence of V and N_o , it suffices by Lemma Appendix B.4 to prove that V and N_o , individually, are Markov. But N_o is Markov since by assumption it has independent increments (this was discussed in section 2) and by invoking Lemma Appendix B.2, (iii). Now to prove that V is Markov, we will show that the assumptions specified by Lemma Appendix B.2 hold and also that condition (iii) in Lemma Appendix B.2 is true. Hence it now suffices to show that

$$R(\rho, \rho') \neq 0 \quad \text{for all } \rho, \rho' \quad (\text{B} - 26)$$

and that (B-4) holds. It is easy to see that the autocorrelation function for V , R_V , is given by

$$R_V(\rho, \rho') = \sigma_c^2 \int h\left(\frac{y}{\rho}\right) h^*\left(\frac{y}{\rho'}\right) dy. \quad (\text{B-27})$$

Assume that $h \in \mathcal{H}$, i.e.,

$$h(y_1, y_2) = C |y_1|^{k_1} |y_2|^{k_2} 1_{[\alpha_1, \beta_1]}(y_1) 1_{[\alpha_2, \beta_2]}(y_2) \quad y_1, y_2 \in \mathbb{R} \quad (\text{B-28})$$

where $\alpha_1 < 0 < \beta_1$, $\alpha_2 < 0 < \beta_2$ and $k_1, k_2 \geq 0$. Then by (B-27) and the above form for h , we see that (B-26) is true. Now let $\rho_l \leq s < t < u \leq \rho_u$. Then

$$\begin{aligned} R_V(u, s) &= \int h\left(\frac{y}{u}\right) h^*\left(\frac{y}{s}\right) dy \\ &= \frac{\sigma_c^2 C^2}{(us)^{k_1+k_2}} \int |y_1|^{2k_1} |y_2|^{2k_2} 1_{[\alpha_1, \beta_1]}\left(\frac{y_1}{u}\right) 1_{[\alpha_1, \beta_1]}\left(\frac{y_1}{s}\right) 1_{[\alpha_2, \beta_2]}\left(\frac{y_2}{s}\right) 1_{[\alpha_2, \beta_2]}\left(\frac{y_2}{u}\right) dy_1 dy_2 \\ &= \frac{\sigma_c^2 C^2}{(us)^{k_1+k_2}} \prod_{j=1}^2 \int_{s\alpha_j}^{s\beta_j} |y_j|^{2k_j} dy_j \end{aligned} \quad (\text{B-29})$$

Also have

$$\begin{aligned} \frac{R_V(u, t) R_V(t, s)}{R_V(t, t)} &= \frac{\int h\left(\frac{y}{u}\right) h^*\left(\frac{y}{t}\right) dy \int h\left(\frac{y'}{t}\right) h^*\left(\frac{y'}{s}\right) dy'}{\int \left|h\left(\frac{y''}{t}\right)\right|^2 dy''} \\ &= \frac{\sigma_c^2 C^2}{(us)^{k_1+k_2}} \frac{\left(\prod_{j=1}^2 \int_{t\alpha_j}^{t\beta_j} |y_j|^{2k_j} dy_j\right) \left(\prod_{j=1}^2 \int_{s\alpha_j}^{s\beta_j} |y_j|^{2k_j} dy_j\right)}{\prod_{j=1}^2 \int_{t\alpha_j}^{t\beta_j} |y_j|^{2k_j} dy_j} \end{aligned} \quad (\text{B-30})$$

By (B-29) and (B-30), we see that (B-4) is satisfied in this case. We omit the proof for the case that h satisfies $\hat{h} \in \mathcal{H}$. It essentially just uses Parseval's theorem and then invokes the equality just proved. This then proves (a).

To prove (b), since N_o has independent increments and is independent of V , it suffices to prove that V has independent increments. First consider the case where h is as above with $k_1 = k_2 = 0$. Then by Lemma Appendix B.5, to prove the result, it suffices to show that

$$R_V(u, s) = R_V(s, s) \quad u \geq s. \quad (\text{B-31})$$

But this is immediate by inspection of (B-29). Again the case where \hat{h} has the above form is handled by Parseval's theorem.

Remark. The above result shows that in the case where h is a sinc function (i.e., represents the Fourier transform of a square aperture) or h is a rect function, our wavelet transform process, I , is an independent increments Gaussian process. By the proof given above, we have actually shown something slightly stronger. We have shown under the hypothesis assumed in part (b) of the Proposition, that I is a time scaled complex Brownian motion. Specifically there exists a non-decreasing function η from $[\rho_l, \rho_u]$ to $[0, \infty)$ such that $I(\rho)$ has the same distribution as the process $Y(\eta(\rho))$ where Y is a complex Brownian motion. For some desired extensions of results given in this appendix, this assumption of h or its Fourier transform being a rect function may provide us with a nice framework for achieving those extensions. There is a lot known about Brownian motion and its properties, and there is a lot known about doing statistical inference for the case of a signal in Brownian motion. Our comment here is simply that this may be very beneficial for future research.

Appendix C Reproducing Kernel Hilbert Spaces Representation

The purpose of this Appendix is to give some background on Reproducing Kernel Hilbert Spaces and how they can be used to represent arbitrary second order stochastic processes. This representation turns out to simplify and make transparent many arguments (even though at first glance this might not appear to be true). This representation is used several places in the main body of the appendix and also is used in the next appendix where we discuss and verify some properties of the generalized signal-to-noise ratios stated in subsections A.1 and A.2.

Background. We suppose that $X = \{X_t : t \in I\}$ is a second order stochastic process with mean 0 and covariance function

$$R(s, t) = E(X(s)X^*(t)) \quad s, t \in I \quad (C-1)$$

where I is an interval in \mathbf{R} . Note we are allowing for the possibility that X is \mathbb{C} -valued. We actually discuss two representations; the Karhunen-Loeve expansion (K-L expansion) and the reproducing kernel Hilbert space representation (RKHS representation). We first make some notes with regards to these representations.

(a) The K-L expansion is applicable when

$$\int \int (R(t, s))^2 dt ds < \infty. \quad (C-2)$$

(b) The RKHS approach is applicable in general, but in these notes we only give this representation under the conditions invoked for the K-L expansion.

(c) All of these representations induce an isomorphism between the Hilbert space generated by linear combinations of $\{X(t) : t \in T\}$, \mathcal{H}_X , and some other Hilbert space. One of the main purposes of doing this is that it allows one to carry out certain projection operations in \mathcal{H}_X by trying to carry them out in the other Hilbert space

(where they may be much easier and/or transparent) and transfer them back via the isomorphism.

(d) Much of the material in these notes are contained in Grenander (1980) and Wahba (1990).

C.1 Karhunen-Loeve Expansion.

Here we suppose that R is continuous and that

$$\int \int (R(s, t))^2 ds dt < \infty. \quad (\text{C} - 3)$$

Then it is easy to show that the linear operator of

$$\phi \in L^2 \rightarrow \int R(s, \cdot) \phi(s) ds \in L^2 \quad (\text{C} - 4)$$

is bounded and compact. Thus by Mercers Theorem (cf. Grenander's *Abstract Inference*), we see that there exists continuous orthonormal functions ϕ_1, ϕ_2, \dots and eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ such that

$$\int R(s, t) \phi_k(s) ds = \lambda_k \phi_k(t) \quad k \in \mathbf{N}, t \in \mathbf{R} \quad (\text{C} - 5)$$

$$R(s, t) = \sum_k \lambda_k \phi_k(s) \phi_k^*(t) \quad s, t \in \mathbf{R} \quad (\text{C} - 6)$$

where the convergence is uniform over compact sets and

$$\int \int (R(s, t))^2 ds dt = \sum_k \lambda_k^2. \quad (\text{C} - 7)$$

Now let

$$Z_k = \sqrt{\lambda_k} \int X(t) \phi_k(t) dt \quad k \in \mathbf{N}, \quad (\text{C} - 8)$$

where the integral exists in the L^2 sense. Then using the above conclusions of Mercer's Theorem, we see that Z_1, Z_2, \dots is a orthonormal sequence of random variables and

$$X(t) = \sum_{k=1}^{\infty} \sqrt{\lambda_k} Z_k \phi_k(t) \quad t \in I \quad (\text{C} - 9)$$

where the sum on the RHS exists in the L^2 sense.

Remarks. (a) It is easy to show that $\mathcal{H}_X \subset \mathcal{H}_Z$ and in the case that $\{\phi_k\}$ forms a complete orthonormal basis then we have equality. Thus in this latter case, it turns out that we have found a countable complete orthonormal basis for \mathcal{H}_X , namely $\{Z_k\}$.

(b) We have to find those ϕ_k 's by solving the integral equation.

(c) Under the conditions specified in (a), we see that we have an isomorphism with ℓ^2 , since $\mathcal{H}_X = \{\sum_k \alpha_k Z_k : \alpha \in \ell^2\}$ and so the isomorphism map is

$$\alpha \in \ell^2 \rightarrow \sum_k \alpha_k Z_k. \quad (\text{C} - 10)$$

C.2 Reproducing Kernel Hilbert Spaces

Here we again suppose the same framework as in the K-L expansion. Now we let

$$\mathcal{H}_1 = \{f \in L^2 : \sum_k \frac{|\langle f, \phi_k \rangle|^2}{\lambda_k} < \infty\} \quad (\text{C} - 11)$$

For $f, g \in \mathcal{H}_1$, we define the inner product by

$$\langle f, g \rangle_1 = \sum_k \frac{\langle f, \phi_k \rangle \overline{\langle g, \phi_k \rangle}}{\lambda_k}. \quad (\text{C} - 12)$$

Proposition. (i) $(\mathcal{H}_1, \langle \cdot, \cdot \rangle_1)$ is a Hilbert Space.

(ii) $f \in \mathcal{H}_1$ implies f is continuous.

(iii) If $f \in \mathcal{H}_1$, then

$$\langle f, R(t, \cdot) \rangle = f(t) \quad f \in \mathcal{H}_1, t \in \mathbb{R}. \quad (\text{C} - 13)$$

(iv) \mathcal{H}_1 and \mathcal{H}_X are isomorphic under the extension of the mapping

$$\sum_{k=1}^K a_k X(t_k) \rightarrow \sum_{k=1}^K a_k R(t_k, \cdot). \quad (\text{C} - 14)$$

Remark. The above is essentially proved in Wahba (1990). But the verification of (iii) is quite easy. In particular

$$\begin{aligned} \langle f, R(t, \cdot) \rangle &= \sum_{k=1}^{\infty} \frac{\langle f, \phi_k \rangle \overline{\langle R(t, \cdot), \phi_k \rangle}}{\lambda_k} \\ &= \sum_{k=1}^{\infty} \frac{\langle f, \phi_k \rangle \sum_{j=1}^{\infty} \lambda_j \phi_j(t) \langle \phi_j, \phi_k \rangle}{\lambda_k} \\ &= \sum_{k=1}^{\infty} \langle f, \phi_k \rangle \phi_k(t) \\ &= f(t). \end{aligned}$$

Also to support the isomorphism statement, we note that

$$\begin{aligned} \langle X(s), X(t) \rangle &= R(s, t) \\ &= \langle R(s, \cdot), R(t, \cdot) \rangle \end{aligned}$$

and we then can use the linearity/continuity properties of the inner products to show the isomorphism.

Remark. One of the main reasons to set up the RKHS approach can be motivated by trying to set up another Hilbert space isomorphism and see why it is not the right embedding. Specifically suppose we tried to set up an identification with L^2 by taking

$$f \in L^2 \rightarrow \int f(t) X(t) dt \quad (\text{C} - 15)$$

and associating the inner product of

$$\langle f, g \rangle_2 = \int \int f(s)g(t)R(s, t) ds dt. \quad (\text{C} - 16)$$

This last expression is truly an inner product and it is true that

$$E \left(\left(\int f(t)X(t) dt \right) \left(\int g(t)X(t) dt \right) \right) = \langle f, g \rangle_2. \quad (\text{C} - 17)$$

We should note that \langle, \rangle_2 is truly an inner product modulo one small detail that it may be possible that for $\langle f, f \rangle_2 = 0$ without f being 0. This small discrepancy is insignificant. However there is another detail about this identification which is not. Specifically this identification will not exhaust \mathcal{H}_X , i.e., not every element in \mathcal{H}_X can be represented by a stochastic integral given by $\int f(t)X(t) dt$ for $f \in L^2$. The reason for this is intuitive by the following argument. Let Z_1, Z_2, \dots be the orthonormal sequence as given in K-L expansion and let $\alpha \in \ell^2$ be such that

$$\sum_k \frac{\alpha_k^2}{\lambda_k} = \infty. \quad (\text{C} - 18)$$

Since $\lambda_k \rightarrow 0$, this is always possible. It will turn out that there is no way to represent $Z \in \mathcal{H}_X$ by an integral $\int f(t)X(t) dt$ since

$$\sum_{k=1}^{\infty} \alpha_k Z_k = \sum_{k=1}^{\infty} \alpha_k \int \frac{\phi_k(t)}{\sqrt{\lambda_k}} X(t) dt. \quad (\text{C} - 19)$$

But we see that $\sum_{k=1}^{\infty} \frac{\alpha_k \phi_k}{\sqrt{\lambda_k}}$ is not in L^2 .

Remark. Note that in general, except for degenerate situations, \mathcal{H}_X is not equal to $L^2(\sigma(X))$, i.e., \mathcal{H}_X is not equal to the space of functions $\psi(X)$ which are square integrable functions ($E(\psi^2(X)) < \infty$). This is easy to see even in the Gaussian case. Specifically it is not too hard to show that all the random variables in \mathcal{H}_X are Gaussian, and this is clearly not the case for $L^2(\sigma(X))$ since $X^2(t)$ is not Gaussian if $X(t)$ is non-degenerate (it has a scaled chi-square distribution), but $X^2(t)$ is

square-integrable. We know that the best estimate of a random variable Y based on knowledge of X is given by $E(Y|X)$ and in general this may not actually be in \mathcal{H}_X . However for the Gaussian case, it is and this is one of the attractive features of the model of Gaussian processes.

Remark. One assumption in Cambanis/Masry (1983) is that the signal function S satisfies that there exists square integrable function f such that

$$\int R(t, s) f(s) ds = S(t). \quad (\text{C} - 20)$$

But this implies that

$$S(\cdot) = \sum_{k=1}^{\infty} \lambda_k \langle f, \phi_k \rangle \phi_k \quad (\text{C} - 21)$$

so that $S = \sum_k \alpha_k \phi_k$ where $\{\frac{\alpha_k}{\sqrt{\lambda_k}}\}_k$ is in ℓ^2 . This allows for the log-likelihood ratio statistic based on knowing all of X and after ignoring terms not dependent on the data X , to be given by

$$\Re \left(\int X(t) f^*(t) dt \right). \quad (\text{C} - 22)$$

Appendix D Properties of GSNRs

In this section we discuss in greater detail some optimality results mentioned in subsections A.1 and A.2, and which are related to choosing resolutions maximizing generalized signal-to-noise ratios (GSNR). These results were used to justify using the criteria of finding resolutions which maximized the generalized signal-to-noise ratio (GSNR). Essentially we are quoting results from Cambanis/Masry (1983), hereafter referred to as CM and also the norm used in this appendix refers to the norm in the reproducing kernel Hilbert space as defined in the previous appendix. In fact in this appendix, we will use all the notation which was set up in the previous appendix, especially the notation for projections. In CM (1983), it was shown that under the assumptions given in (a1) through (a5), have for any set of resolutions represented by \mathcal{L}_n , that the corresponding probability of detection is given by

$$P_D(\mathcal{L}_n) = \Phi(\|\mathcal{P}_{\mathbf{L}_n} S\| - \Phi^{-1}(1 - P_{FA})) \quad (D-1)$$

where \mathbf{L}_n is linear subspace generated in the reproducing kernel Hilbert space by the functions $\{R(\rho_j, \cdot) : 1 \leq j \leq n\}$, ϕ is probability density function of standard real Gaussian, i.e.,

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad z \in \mathbf{R} \quad (D-2)$$

and Φ is the cumulative distribution function for the above probability density function, i.e.,

$$\Phi(z) = \int_{-\infty}^z \phi(z') dz' \quad z \in \mathbf{R}. \quad (D-3)$$

Note that this implies that the sequence of resolutions which maximize the GSNR ($=\|\mathcal{P}_{\mathbf{L}_n} S\|$) also maximize the probability of detection $P_D(\mathcal{L}_n)$. Hence the justification for focusing on the resolutions which maximize GSNR. Based on the above result in (D-1), it is an straightforward exercise in calculus to show that if \mathcal{L}_n^o is an optimal sequence for maximizing probability of detection or equivalently maximizing

the GSNR, then

$$\frac{P_D^\dagger - P_D^n}{\|S\|^2 - \|\mathcal{P}_{\mathbf{L}_n} S\|^2} \rightarrow 2\|S\|^2 \phi \left[\|S\|^2 - \Phi^{-1}(1 - P_{FA}) \right]. \quad (\text{D} - 4)$$

This result can be used to help choose the constant K in the general loss function given in subsection A.2. Also it can be used to derive alternative algorithms for finding resolutions without directly trying to maximize GSNR. This was discussed in some detail in subsection A.1 for the case where the noise process had independent increments. Based on (D-4), it is straightforward to define the notion of a sequence \mathcal{L}_n^1 being asymptotically optimal. Specifically let \mathcal{L}_n^o be a sequence of optimal resolutions. We say a sequence of vector of resolutions \mathcal{L}_n^1 is asymptotically optimal if and only if

$$\frac{\|S\|^2 - \|\mathcal{P}_{\mathbf{L}_n^1} S\|^2}{\|S\|^2 - \|\mathcal{P}_{\mathbf{L}_n^o} S\|^2} \rightarrow 1 \quad (\text{D} - 5)$$

where \mathbf{L}_n^o represents the linear space spanned by $\{R(\rho_j^o, \cdot) : 1 \leq j \leq n\}$ and \mathbf{L}_n^1 represents the linear space spanned by $\{R(\rho_j^1, \cdot) : 1 \leq j \leq n\}$. For a further discussion of these and related issues see CM (1983).

References

- [1] Beckman P., and Spizzichino, A., *The Scattering of Electromagnetic Waves From Rough Surfaces*, Artech House, Norwood, MA (1987).
- [2] Walker, J. L., "Range-Doppler Imaging of Rotating Objects," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES- 16, p. 23, Jan. 1980.
- [3] Flandrin, P., "Wavelets and Related Time-Scale Transforms," in: *Advanced Signal Processing Algorithms, Architectures, and Implementations*, F.T. Luk (ed.), Proc. SPIE, vol. 1348, 1990.
- [4] VanTrees, H.L., *Detection, Estimation, and Modulation Theory—Part 1*, Wiley, New York, (1968).
- [5] Wong, E., and Hajek, B., *Stochastic Processes in Engineering Systems*, Springer, Berlin, (1985).
- [6] Fukunaga, K., and Hayes, R.R., "Effects of Sample Size in Classifier Design," *IEEE Trans. PAMI*, 11, No. 8, August 1989.
- [7] Papoulis, A., *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, New York, (1965).
- [8] Duda, R. O., Hart P. E., *Pattern Classification and Scene Analysis*, Wiley, New York, (1973).
- [9] Anderson, T.W., *An Introduction to Multivariate Statistical Analysis, 2nd Edition*, Wiley, New York, (1971).
- [10] Cambanis, S. and Masry, E. (1983). Sampling designs for the detection of signals in noise, *IEEE Trans. on IT*, vol. IT-29, 83-104.
- [11] Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets, *Comm. on Pure and Applied Math.*, vol. XLI, pp. 909-996.

- [12] Grenander, U. (1980). *Abstract Inference*, Wiley, New York.
- [13] Heil, C. and Walnut, D. (1989). Continuous and Discrete Wavelet Transforms, *SIAM Review*, vol. 31, pp. 628-666.
- [14] Strang, G. (1989). Wavelets and dilation equations; a brief introduction, *SIAM Review*, vol. 31, pp. 614-627.
- [15] Wahba, G. (1990). *Spline Models for Observational Data*, CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia.
- [16] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis, 2nd Edition*, Wiley, 1984.
- [17] S. F. Arnold, *The Theory of Linear Models and Multivariate Analysis*, Wiley, 1981.
- [18] P. J. Brockwell and R. A Davis, *Time Series: Theory and Methods*, Springer, 1987.
- [19] S. R. DeGraaf, "Sidelobe Reduction via FIR Filtering in SAR Imagery," *IEEE Trans. Image Proc.*, vol 3, no. 3, 1994.
- [20] E. J. Kelly, "An Adaptive Detection Algorithm," *IEEE T Aero. Elec. Sys.*, AES-22, 1, 1986.
- [21] G. Minler and J. Minkler, *CFAR: The Principles of Automatic Radar Detection in Clutter*, Magellan, Baltimore, 1990.
- [22] S. R. Sullivan, et. al., "Strategic Target Algorithm Research (STAR) - Final Report," Contract Number F19628-90-C-0002, 1991, SECRET.
- [23] B. J. Thelen, "Complete Characterization of Complex Distributions from Weak Scatterers," *J. Opt. Soc. Amer., ser. A*, in review, 1994.

- [24] D. Snyder, *Random Point Processes*, Wiley, 1975.
- [25] Orphanidis, S., *Optimum Signal Processing*, MacMillan, New York, (1988).
- [26] S. Cambanis and E. Masry, "Sampling Designs for the Detection of Signals in Noise," *IEEE T Info. Theory*, IT-29, 1, 1983.
- [27] Y. Meyer, *Wavelets: Algorithms and Applications*, SIAM , Philadelphia, 1993.
- [28] W. L. Root, "The detection of signals in Gaussian noise," in *Communication Theory*, A.V. Balakrishnan, ed., McGraw-Hill, New York, 1968.
- [29] L. Breiman, *et. al.*, *Classification and Regression Trees*, (Wadsworth Press, Pacific Grove, CA), 1984.
- [30] G.H. Ball and D.J. Hall, "A clustering technique for summarizing multivariate data," *Behavioral Sci.*, vol. 12, pp. 153-155, March 1967.
- [31] P.A. Chou, "Optimal Partitioning for Classification and Regression Trees," *IEEE Trans. Pattern Anal. and Mach. Intel.*, vol. 13, pp. 340-353, April 1991.
- [32] L. M. Novak, M. C. Burl, R. D. Chaney, G. J. Owirka, "Optimal Processing of Polarimetric Synthetic-Aperture Radar Imagery", *The Lincoln Laboratory Journal*, vol. 3, no. 2, 1990.
- [33] J. Jao, "Amplitude Distribution of Composite Terrain Clutter and the K-Distribution", *IEEE Trans. Ant. Prop.*, vol. AP-32, no. 10, Oct. 1984.